

ANTWORTSTABILE
REKURSIVE DIGITALFILTER
IN FESTKOMMA-ARITHMETIK

von
Roland Czarnach

Erlangen, 1984

ANTWORTSTABILE REKURSIVE DIGITALFILTER IN FESTKOMMA-ARITHMETIK

Der Technischen Fakultät
der Universität Erlangen - Nürnberg
zur Erlangung des Grades
DOKTOR - INGENIEUR

vorgelegt von
Roland Czarnach

Erlangen, 1984

Als Dissertation genehmigt von der
Technischen Fakultät der Universität Erlangen-Nürnberg.

Tag der Einreichung: 14. Februar 1984

Tag der Promotion: 2. August 1984

Dekan: Prof. G. Vetter

Berichterstatter: Prof. Dr.-Ing. W. Schüssler

Prof. Dr.-Ing. R. Unbehauen

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. W. Schüssler für die interessante Aufgabenstellung, die zahlreichen Anregungen und Hinweise und seine wertvolle Hilfe bei der Abfassung der Niederschrift.

Herrn Prof. Dr.-Ing. R. Unbehauen danke ich für die Übernahme des Zweitreferates und sein damit bekundetes Interesse an der Arbeit.



Digitized by the Internet Archive
in 2026

I N H A L T S V E R Z E I C H N I S

Seite

LISTE DER WICHTIGSTEN FORMELZEICHEN

1. PROBLEMSTELLUNG UND ÜBERSICHT	1
2. THEORIE ÜBERLAUFSTABILER FILTERSTRUKTUREN	5
2.1 Vorbemerkungen	5
2.2 Analyse mit Hilfe der Zustandsraumbeschreibung	7
2.2.1 Zustandsraumbeschreibung	7
2.2.2 Grenzzzyklusfreiheit	9
2.2.3 Antwortstabilität	18
2.2.4 Ergänzungen	22
2.3 Analyse mit Hilfe eines verallgemeinerten Modells	23
2.3.1 Modell-Beschreibung	23
2.3.2 Grenzzzyklusfreiheit	26
2.3.3 Antwortstabilität	31
2.3.4 Zusammenhang mit der $\underline{A}$ -Matrix	36
2.3.4 Modifikation des allgemeinen Modells	40
2.4 Zusammenfassung	42
3. ÜBERLAUFSTABILE STRUKTUREN - BESCHREIBUNG UND VERGLEICHENDE UNTERSUCHUNGEN	44
3.1 Vorbemerkungen	44
3.2 Zustandsraum-Strukturen	45
3.2.1 Aussteuerung und Rundungsrauschen	45
3.2.2 Normalform-Struktur	45
3.2.3 Rauschoptimale Struktur (Mullis/Roberts)	49
3.2.3.1 Vorbemerkung	49
3.2.3.2 Filterblock 2ten Grades	50

3.2.4 Zusammenfassung	58
3.3 Allgemeine Wellendigitalfilter	59
3.3.1 Vorbemerkungen	59
3.3.2 Netzwerkbeschreibung durch Wellengrößen	60
3.3.3 Zusammenhang s- und z-Bereich	61
3.3.4 Wellenbeziehungen elementarer Bausteine	62
3.3.5 Verbindung der Tore zum Netzwerk	65
3.3.6 Allgemeine Anmerkungen zur Realisierung	66
3.3.7 Überlaufstabilität	67
3.3.8 Eigenschaften der $\underline{A}$ -Matrix	70
3.3.8.1 Allgemeine Wellendigitalfilter	70
3.3.8.2 R, L, C, Ü - Referenzfilter	77
3.3.8.3 RLÜ- oder RCÜ-Referenzfilter	82
3.4 Leiterstrukturen nach Markel/Gray	83
3.4.1 Beschreibung der Strukturen	83
3.4.2 Skalierung	87
3.4.3 Überlaufstabilität	88
3.4.4 Systemmatrix $\underline{A}$	91
4. ZUSAMMENFASSUNG UND ERGÄNZUNG DER BISHERIGEN ERGEBNISSE	96
4.1 Modelle zur Analyse der Überlaufstabilität	96
4.2 Eigenschaften von $\underline{A}$, Vergleich der Filterstrukturen	99
5. STRUKTURVERGLEICH BEZÜGLICH DES RUNDUNGSRAUSCHENS UND DER GRANULAREN GRENZZYKLEN	107
5.1 Vorbemerkungen	107
5.2 Untersuchte Strukturen	108
5.3 Rundungsrauschen	111
5.4 Granulare Grenzzyklen	125

III

	Seite
5.4.1 Meßergebnisse	125
5.4.2 Theoretische Abschätzung der Grenzzyklus- Amplituden	134
6. FILTER-REALISIERUNG DURCH BLOCKVERARBEITUNG	142
6.1 Vorbemerkungen	142
6.2 Der Filter-Algorithmus	143
6.3 Quantisierungseigenschaften	145
6.3.1 Rundungsrauschen	145
6.3.2 Überlaufstabilität	147
6.3.3 Granulare Grenzzyklen	149
6.4 Meßergebnisse	150
7. STABILITÄT ZEITVARIABLER DIGITALER FILTER	154
7.1 Vorbemerkungen	154
7.2 Stabilität eines Filters zweiten Grades am Beispiel der ersten kanonischen Form	155
7.3 Analytische Beschreibung periodisch zeit- varianter Systeme	159
7.4 Stabilität durch Wahl einer geeigneten Filter- struktur	161
7.5 Äquivalente Realisierungen mit unterschied- lichen Filterstrukturen	163
A N H A N G	166
A 1	166
A 2	171
A 3	177
A 4	186
A 5	196
A 6	201
A 7	205
L I T E R A T U R	207

LISTE DER WICHTIGSTEN FORMELZEICHEN ---

$\underline{A}, \underline{B}, \underline{C}, \underline{D}$	Koeffizientenmatrizen der Zustandsraum-Beschreibung
$\underline{a}(k)$	Eingangsvektor (allgemeines Filtermodell)
$\underline{b}(k)$	Ausgangsvektor (allgemeines Filtermodell)
$\underline{E}$	Einheitsmatrix
$\underline{G}$	Diagonalmatrix mit positiv reellen Elementen
$H(z), F(z)$	Übertragungsfunktionen
j	Imaginäre Einheit
$\underline{K}$	Kovarianzmatrix der Impulsantworten vom Filtereingang zu allen Zustandsvariablen
k	Laufindex einer Signalfolge
n	Filtergrad
$\boxed{Q}$	Quantisierungsintervall
Q	Quantisierer einschließlich Überlaufkennlinie
R_i	Innere Rauschkennzahl
$\underline{S}$	Streumatrix
$\boxed{T}$	Zustandsspeicher
$\underline{U}$	allgemeines Symbol für unitäre oder orthogonale Matrizen
$\underline{V}, \underline{V}_*$	Verknüpfungsmatrizen des allgemeinen Filtermodells
$v(k)$	Filter-Eingangsfolgenvektor
$y(k)$	Filter-Ausgangsfolgenvektor

λ_v	Eigenwert
$\underline{\Lambda}$	Diagonalmatrix der Eigenwerte
ρ	Polradius
ψ	Polwinkel
σ^2	Varianz des Rundungsrauschens
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{R}^n$	n-dim. reeller Vektorraum
$\forall$	Symbol "für alle"
$ \dots $	Norm
$ \dots _{\underline{G}}$	Vektornorm, definiert mittels einer Diagonalmatrix, z.B. $ \underline{x} _{\underline{G}} = \sqrt{\underline{x}^T \underline{G} \underline{x}}$

1. PROBLEMSTELLUNG UND ÜBERSICHT

Bei der Realisierung digitaler Filter, insbesondere der rekursiven Systeme, treten infolge der begrenzten Wortlänge eine Reihe numerischer Probleme (Quantisierungseffekte) auf. Beschränken wir uns auf den üblichen Fall der Verarbeitung in Festkomma-Arithmetik, so sind folgende Fehlerarten zu beachten:

Koeffizientenfehler:

Sie führen zu einer Verfälschung des Übertragungsverhaltens, sollen hier aber nicht näher diskutiert werden.

Rundungsfehler:

Nach Multiplikationen ist i.a. eine Verkürzung der Wortlänge auf die ursprüngliche Stellenzahl erforderlich. Bei stochastischen Filtereingangssignalen führt dies zu einem überlagerten Störgeräusch am Ausgang. Bei verschwindender oder periodischer Erregung können parasitäre Oszillationen (granulare Grenzzyklen) auftreten.

Überlauffehler:

Bei Übersteuerung des begrenzten Zahlenbereichs (Intervall $[-1, +1]$) entstehen relativ große nichtlineare Fehler. Sie können bei anschließend verschwindender Erregung sog. Überlaufgrenzzyklen hoher Amplitude auslösen. Bei stochastischer Erregung sind permanente Störungen möglich, die das Ausgangssignal völlig verfälschen; dies gilt auch dann, wenn im idealen linearen Falle nur selten Zahlenbereichsüberschreitungen auftreten würden.

Derartige Effekte sind somit als besonders kritisch anzusehen.

Zur Vermeidung von Überlaufgrenzzyklen wurden bereits 1969 von Ebert, Mazo und Taylor /12/ spezielle Überlauf-Charakteristiken (z.B. Sättigungs-, Dreieckskennlinie) vorgeschlagen; die Ergebnisse gelten für Filterblöcke zweiten Grades in weitgehend beliebiger Struktur. Diese Maßnahme gewährleistet i.a. jedoch nicht die Antwortstabilität bei Erregung (forced response stability) /7/.

Einen Zusammenhang zwischen der Freiheit von Überlaufgrenzzyklen und der Antwortstabilität eines Systems stellten erstmals Claasen, Mecklenbräuker und Peek /8/ her. Dabei zeigte sich, daß die von Fettweis /2/ vorgeschlagenen Wellendigitalfilter bei Verwendung der genannten Kennlinien antwortstabil sind.

In jüngerer Zeit wurden spezielle "Zustandsraum-Strukturen" /13, 21/ vorgeschlagen, welche bei beliebigen Überlauf-Charakteristiken grenzzyklusfrei sind.

Im ersten Teil der vorliegenden Arbeit soll eine Vereinheitlichung der unterschiedlichen Ansätze zum Stabilitätsnachweis erarbeitet werden. Ferner sollen Gemeinsamkeiten und Unterscheidungsmerkmale der antwortstabilen Strukturen herausgestellt werden, was letztlich in eine Klassifizierung bezüglich des Stabilitätsverhaltens und der Eigenschaften der Systemmatrix $\underline{A}$ (Zustandsraum-Beschreibung) mündet.

Dazu wird in Kapitel 2 zunächst das bekannte Kriterium /3/ für die Grenzzyklusfreiheit von Zustandsraum-Strukturen erläutert und insbesondere auch der Bezug zur Antwortstabilität hergestellt. Es wird ausführlich hergeleitet, daß die im Sinne des genannten Kriteriums grenzzyklusfreien Filter ebenfalls antwortstabil sind, sofern die in /12/ vorgeschlagenen Überlauf-Kennlinien realisiert werden. Es handelt sich hierbei um eine besondere Form der Antwortstabilität in dem Sinn, daß die Störung mit jedem Filterschritt abklingt. Insofern geht sie über die in /8/ formulierten Forderungen hinaus.

Im nächsten Schritt wird ein allgemeines Filtermodell zugrundegelegt und hierzu ein neues Stabilitätskriterium eingeführt. Es stellt bei beliebigen Filterstrukturen eine hinreichende Bedingung für die Antwortstabilität dar. Im Fall der Zustandsraumstrukturen können damit neue Eigenschaften angegeben werden, wie sie bisher nur für Wellendigitalfilter bekannt waren und unter dem Begriff "inkrementale Pseudopassivität" zusammengefaßt wurden /9/.

In Kapitel 3 werden die bekanntesten antwortstabilen Strukturen gegenübergestellt, wobei u.a. die Anwendung des neu eingeführten Stabilitätskriteriums demonstriert wird.

Die wichtigsten theoretischen Ergebnisse sind schließlich in Kapitel 4 noch einmal übersichtlich zusammengefaßt.

Kapitel 5 enthält einen Strukturvergleich bezüglich des Rundungsrauschens und der granularen Grenzzyklen. Die Resultate wurden theoretisch und durch Messung ermittelt.

In diesem Zusammenhang erfolgt auch eine theoretische Abschätzung der Amplituden granularer Grenzzyklen, was eine Ergänzung der bekannten z.B. in /51/ zusammengefaßten Ergebnisse darstellt.

Als besonders günstig erweist sich die aus rauschoptimalen Filterblöcken zweiten Grades nach Mullis und Roberts /21/ aufgebaute Struktur, deren Nachteil aber in einer vergleichsweise hohen Multipliziererschrittzahl besteht.

Eine Verbesserung bringt hier die z.B. von Zeman und Lindgren /40/ vorgeschlagene Blockverarbeitung. Damit ist neben der Reduzierung des arithmetischen Aufwandes auch eine weitere Verminderung des Rundungsrauschens sowie der granularen Grenzzyklen verbunden.

In Kapitel 6 wird das Verfahren erläutert und durch eine verallgemeinerte theoretische Analyse der Überlaufstabilität sowie der granularen Grenzzyklen erläutert. Abschließend werden einige Meßergebnisse vorgestellt.

Die theoretischen Überlegungen zur Überlaufstabilität regten zur Diskussion eines Nebenaspektes, der Stabilität zeitvariabler Digitalfilter an (Kapitel 7). Obwohl das Problem eigentlich etwas außerhalb der bisherigen Thematik liegt, rechtfertigen doch die bei der Betrachtung des Zustandsraum-Verhaltens gewonnenen Einblicke diesen ergänzenden Abschnitt.

Das Stabilitätsverhalten wird zunächst anschaulich erklärt und anhand von Beispielen illustriert. Sodann wird gezeigt, wie durch Wahl geeigneter Filterstrukturen die Stabilität stets gesichert werden kann. Schließlich wird eine Transformation angegeben, bei deren Anwendung das Systemverhalten unabhängig von der Filterstruktur ist.

Insgesamt ist mit der vorliegenden Arbeit eine zusammenfassende Darstellung antwortstabiler Filterstrukturen gegeben, die neben einer verallgemeinerten Theorie die bekannten Ergebnisse in vielen Punkten ergänzt bzw. erweitert. Die Ergebnisse aus den theoretischen und praktischen Untersuchungen des Rundungsrauschens und der granularen Grenzyklen sollen schließlich dem Anwender Kriterien für die Auswahl geeigneter Filterstrukturen liefern.

2. THEORIE ÜBERLAUFSTABILER FILTERSTRUKTUREN

2.1 Vorbemerkungen

Das Ziel des vorliegenden Kapitels ist es, eine möglichst allgemeingültige Beschreibung antwortstabiler Digitalfilter zu finden, die an späterer Stelle eine Spezialisierung auf die betrachteten Strukturen erlaubt. Diese lassen sich in zwei Gruppen einteilen, die Zustandsraum-Strukturen und die Wellendigitalfilter.

Zustandsraum-Strukturen:

sind dadurch gekennzeichnet, daß der Verarbeitungsalgorithmus unmittelbar der Zustandsraumbeschreibung entspricht. Insbesondere ist jedem Element der Systemmatrizen A, B, C, D ein Multiplizierer in der Realisierung zugeordnet. Eine genauere Darstellung sowie das Beispiel eines derartigen Filterblockes zweiten Grades findet sich im folgenden Abschnitt 2.2.1.

Wellendigitalfilter:

sind von analogen Netzwerken abgeleitet /2/, wobei den einzelnen Elementen ein digitales Äquivalent zugeordnet wird. Die Verbindung erfolgt über sogenannte "Adaptoren". Eine detaillierte Beschreibung dieser Strukturen wird an späterer Stelle gegeben.

Die Analyse des Grenzzyklusverhaltens sowie der Antwortstabilität kann für die erste Gruppe in sehr anschaulicher Weise auf der Basis der Zustandsraumbeschreibung durchgeführt werden, was in Abschnitt 2.2 dargelegt wird. Im Fall der Wellendigitalfilter wurde dagegen mit Hilfe der Begriffe "Pseudoleistung" und "Pseudopassivität" bislang ein grundsätzlich anderer Weg beschritten. Um beide Ansätze zu vereinheitlichen, wird in Abschnitt 2.3 ein verallgemeinertes Filtermodell zugrunde gelegt. Es werden hinreichende Bedingungen für Antwortstabilität und Grenzzyklusfreiheit formuliert, die sowohl von den interessierenden Zustandsraumfiltern

als auch von den Wellendigitalfiltern erfüllt werden. Damit können insbesondere im Fall der Zustandsraumstrukturen neue Eigenschaften angegeben werden, wie sie bisher nur für Wellendigitalfilter bekannt waren und unter der Bezeichnung "inkrementale Pseudopassivität" zusammengefaßt wurden.

Zur Vermeidung von Mißverständnissen sollen einige Begriffe nochmals wie folgt vereinbart werden:

Überlaufgrenzzyklen:

periodische Schwingung, ausgelöst durch Überschreiten des zulässigen Zahlenbereiches (= Überlauf), die trotz verschwindender Erregung und Stabilität im linearen Fall (Polstellen im Innern des Einheitskreises) nicht abklingt.

Anwortstabilität (forced response stability):

Eine durch Überlauf hervorgerufene Störung klingt auch bei nicht verschwindender Erregung stets ab, sofern unter idealen linearen Bedingungen Überläufe nur vorübergehend auftreten.

Praktisch bedeutet dies: bei selten auftretenden Überläufen klingt die Störung zwischenzeitlich ab.

Antwortstabilität schließt die Freiheit von Überlaufgrenzzyklen ein.

Inkrementale Pseudopassivität:

Steigerung der Antwortstabilität; bedeutet u.a., daß permanent auftretende geringfügige Überläufe auch nur geringfügige Störungen des Ausgangssignals zur Folge haben.

Überlaufstabilität:

Pauschaler Oberbegriff, der im Einzelfall genauer zu spezifizieren ist.

Granularer Grenzyklus:

Periodische Schwingung, ausgelöst durch Rundungsfehler.

2.2 Analyse mit Hilfe der Zustandsraumbeschreibung

In diesem Abschnitt wird eine zusammenfassende Darstellung der bekannten Theorie überlaufstabiler Zustandsraum-Strukturen gegeben. Die Ausführungen orientieren sich im wesentlichen an den Arbeiten von Mills, Mullis, Roberts /3/ und Claasen, Mecklenbräuer, Peek /8/.

2.2.1 Die Zustandsraumbeschreibung

Ein wichtiges Hilfsmittel zur Beschreibung kausaler rekursiver digitaler Filter ist ihre Darstellung im Zustandsraum /52, 49/ mit Hilfe der Matrix-Beziehungen

$$\begin{aligned}\underline{x}(k+1) &= \underline{A} \underline{x}(k) + \underline{B} \underline{v}(k) \\ \underline{y}(k) &= \underline{C} \underline{x}(k) + \underline{D} \underline{v}(k)\end{aligned}\tag{2-1}$$

mit

$\underline{x}(k)$ = n-dimensionaler Spaltenvektor der Zustandsvariablen

$\underline{v}(k)$ = l-dimensionaler Spaltenvektor der Eingangsfolgen

$\underline{y}(k)$ = r-dimensionaler Spaltenvektor der Ausgangsfolgen

$\underline{A}$ = n×n Matrix

$\underline{B}$ = n×l Matrix

$\underline{C}$ = r×n Matrix

$\underline{D}$ = r×l Matrix

Die Matrizen $\underline{A}$, $\underline{B}$, $\underline{C}$, $\underline{D}$ kennzeichnen die Filtereigenschaften und besitzen im Fall zeitinvarianter Systeme, welche hier ausschließlich betrachtet werden, konstante Koeffizienten.

Der Zusammenhang mit der Matrix $\underline{H}(z)$ der Übertragungsfunktionen von allen Eingängen zu allen Ausgängen ist gegeben durch

$$\underline{H}(z) = \underline{C}[\underline{zE} - \underline{A}]^{-1} \cdot \underline{B} + \underline{D} \quad (2-2)$$

Zur Erläuterung der Zustandsraumdarstellung (2-1) möge ein einfacher Filterblock 2. Grades mit einem Eingang und einem Ausgang dienen. (2-1) läßt sich dann schreiben als

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \cdot v(k) \quad (2-3)$$

$$y(k) = [c_1, c_2] \cdot \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + d \cdot v(k)$$

Bild 2.1 zeigt die zugehörige Realisierung in Zustandsraumstruktur.

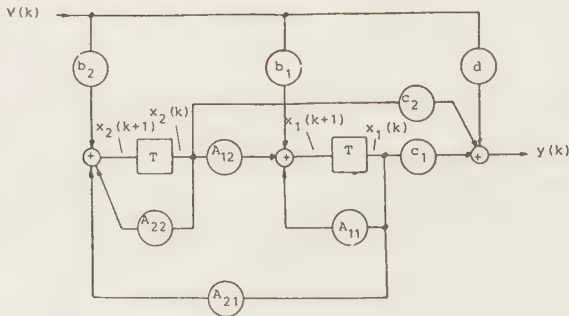


Bild 2.1: Filterblock 2. Grades in Zustandsraumstruktur

Bei gegebenem $\underline{H}(z)$ sind die Matrizen $\underline{A}$, $\underline{B}$, $\underline{C}$, $\underline{D}$ nicht eindeutig bestimmt. So läßt sich der ZV-Vektor (ZV für Zustandsvariable) $\underline{x}(k)$ mittels einer beliebigen nichtsingulären $n \times n$ Matrix $\underline{T}$ in einen geänderten Vektor $\underline{x}'(k) = \underline{T}^{-1} \underline{x}(k)$ transformieren.

Einsetzen von $\underline{x}(k) = \underline{T} \underline{x}'(k)$ in (2-1) liefert dann eine modifizierte Zustandsraumdarstellung

$$\begin{aligned} \underline{x}'(k+1) &= \underbrace{\underline{T}^{-1} \underline{A} \underline{T}}_{= \underline{A}'} \underline{x}'(k) + \underbrace{\underline{T}^{-1} \underline{B}}_{= \underline{B}'} \underline{v}(k) \\ \underline{y}(k) &= \underbrace{\underline{C} \underline{T}}_{= \underline{C}'} \underline{x}'(k) + \underline{D} \underline{v}(k) \end{aligned} \quad (2-4)$$

Die hieraus für die Realisierung resultierenden Freiheitsgrade lassen sich z.B. für den Entwurf rauscharmer und überlaufstabiler Strukturen ausnutzen.

2.2.2 Grenzzykusfreiheit bei verschwindender Erregung

Für die folgenden Überlegungen wird gefordert, daß Quantisierungsfehler infolge der endlichen Wortbreite nur jeweils nach Berechnung der neuen Zustandsvariablen $x_v(k+1)$, $v = 1(1)n$, auftreten können. Für ein Filter 2. Grades in Zustandsraumstruktur ist dies in Bild 2.2 dargestellt, wobei $\boxed{Q}$ die Quantisierung symbolisiert. Um diese Forderung in der Praxis zu gewährleisten, sind die auftretenden Teilprodukte zunächst mit erhöhter Wortbreite zu adaptieren. Bei der anschließenden Wortlängenverkürzung können folgende Fehler auftreten:

- Fehler durch Rundung, Betragsabschneiden u.ä.
- Fehler durch Überschreiten des zulässigen Zahlenbereichs $[-1, +1)$,

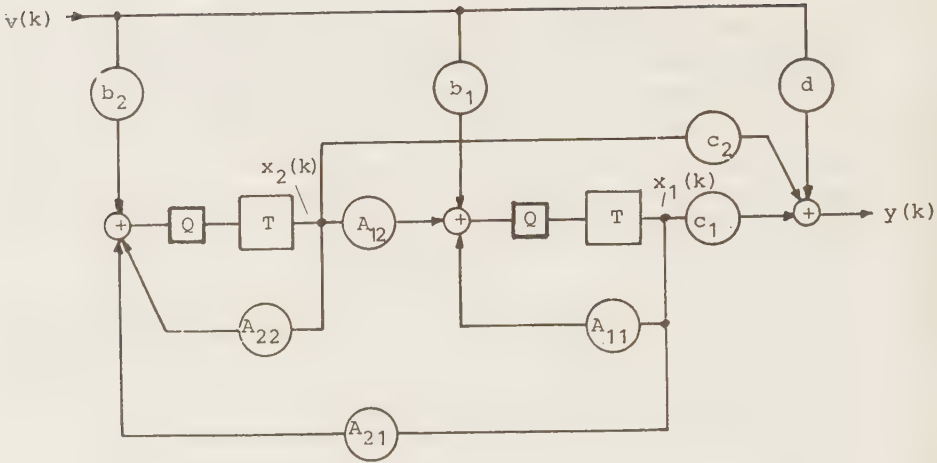


Bild 2.2: Beispiel für Quantisierung nach Zusammenfassen der Teilprodukte

Für Filterkoeffizienten, die im linearen Fall ein stabiles System beschreiben, können die genannten Fehler (Nichtlinearitäten) dennoch zu Grenzyklen führen /6/. Es handelt sich hierbei um periodische Signalfolgen, die trotz verschwindender Erregung nicht abklingen. Wir unterscheiden zwischen granularen Grenzyklen (Fehler a)) und Überlauf-Grenzyklen, deren Ursache in der Begrenzung des Zahlenbereiches (Fehler b)) liegt; sie können sehr hohe Amplituden erreichen. Das vorliegende Kapitel beschäftigt sich zwar primär mit der Überlauf-Stabilität, jedoch besteht ein gewisser Zusammenhang mit der Freiheit von granularen Grenzyklen, so daß diese der Vollständigkeit halber mit behandelt werden.

Im folgenden wird angenommen, daß der Quantisierer einem Eingangswert $\tilde{x}_v$ einen Ausgangswert $x_v = [\tilde{x}_v]_Q$ derart zuordnet, daß stets gilt:

$$\left| \underbrace{[\tilde{x}_v]_Q}_{x_v} \right| \leq |\tilde{x}_v| \quad (2-5)$$

Eine Quantisierer-Kennlinie, welche dieser Bedingung genügt, ist für eine Wortlänge von $w = 3$ Bit in Bild 2.3 dargestellt. Dabei erfolgt im Bereich $-1 < \tilde{x}_v < 1$ die Wortlängenverkürzung durch Betragsabschneiden, während für $|\tilde{x}_v| \geq 1$ eine beliebige Zuordnung innerhalb der schraffierten Fläche zulässig ist.

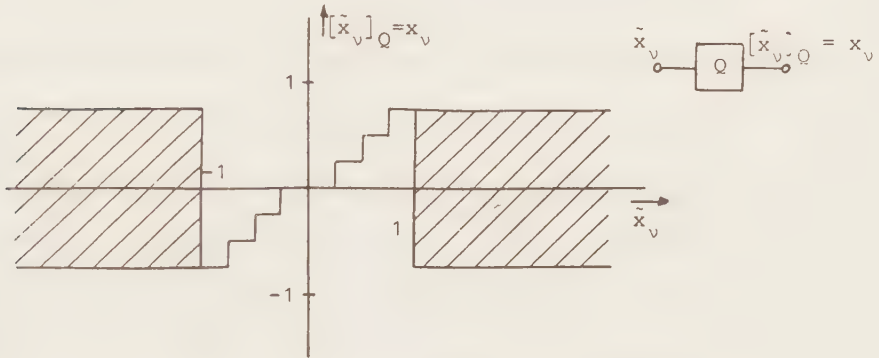


Bild 2.3: Beispiel einer zulässigen Quantisierer-Kennlinie

Eine hinreichende Bedingung für die Grenzzzyklusfreiheit ist nun offensichtlich, daß eine geeignet definierte Norm $||\underline{x}(k)||_{\underline{G}}$ des ZV-Vektors $\underline{x}(k)$ bei verschwindender Erregung in jedem Filterschritt abnimmt. Wir definieren

$$||\underline{x}||_{\underline{G}} = \sqrt{\underline{x}^T \underline{G} \underline{x}}$$

wobei

(2-6)

$$\underline{G} = \text{Diag} [G_1, G_2, \dots, G_n]$$

eine beliebige konstante Diagonalmatrix mit positiv reellen Elementen ist. Mit

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

gilt dann

$$||\underline{x}||_{\underline{G}}^2 = G_1 x_1^2 + G_2 x_2^2 + \dots + G_n x_n^2. \quad (2-7)$$

Gemäß obiger Bedingung soll für beliebige ZV-Vektoren $\underline{x}(k)$ gelten

$$||\underline{x}(k+1)||_{\underline{G}} < ||\underline{x}(k)||_{\underline{G}} \quad \forall \underline{x}(k) \neq \underline{0}. \quad (2-8)$$

Unter Berücksichtigung der Zustandsraumdarstellung (2-1) und der Quantisierung geht (2-8) über in

$$||\underline{x}(k+1)||_{\underline{G}} = ||[\underline{A} \underline{x}(k)]_Q||_{\underline{G}} < ||\underline{x}(k)||_{\underline{G}}. \quad (2-9)$$

Aus den Beziehungen (2-5) und (2-7) folgt ferner

$$||[\underline{A} \underline{x}(k)]_Q||_{\underline{G}} \leq ||\underline{A} \underline{x}(k)||_{\underline{G}}. \quad (2-10)$$

D.h., die Beziehung

$$||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \neq \underline{0} \quad (2-11)$$

stellt eine hinreichende Bedingung für asymptotische Stabilität und damit Grenzzyklusfreiheit dar, wenn die Quantisierer die Voraussetzung (2-5) erfüllen.

Anhand eines Filters 2. Grades soll dies veranschaulicht werden. Für die Norm des ZV-Vektors gilt dann

$$||\underline{x}||_{\underline{G}}^2 = G_1 x_1^2 + G_2 x_2^2. \quad (2-12)$$

Bei konstanter Norm $||\underline{x}||_{\underline{G}}$ beschreibt (2-12) eine achsensymmetrische Ellipse in der x_1 - x_2 -Zustandsebene mit den Halbachsen

$$\frac{||\underline{x}||_{\underline{G}}}{\sqrt{G_1}} \quad \text{und} \quad \frac{||\underline{x}||_{\underline{G}}}{\sqrt{G_2}} \quad (\text{Ellipse I in Bild 2.4})$$

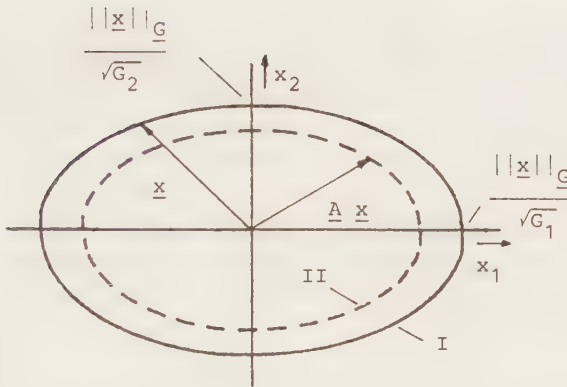


Bild 2.4: Linien konstanter Norm in der Zustandsebene

Wenn nun ein beliebiger ZV-Vektor $\underline{x}(k)$ einen Punkt dieser Ellipse beschreibt, so beschreibt nach einem Filterschritt der neue ZV-Vektor $\underline{x}(k+1)$ im Idealfall, d.h. ohne Berücksichtigung der Quantisierung, wegen der Bedingung (2-11) einen Punkt auf einer kleineren Ellipse mit dem gleichen Halbachsenverhältnis (Ellipse II in Bild 2.4). Da Quantisierungsfehler nach Voraussetzung allenfalls eine betragsmäßige Verkleinerung der entsprechenden Zustandsvariablen bewirken, wird der tatsächlich berechnete ZV-Vektor einen Punkt auf oder im Innern von Ellipse II kennzeichnen.

Offensichtlich konvergiert die Länge des ZV-Vektors mit wachsendem k gegen Null.

Als Realisierungsbeispiel möge ein Filter zweiten Grades in Normalform-Struktur /1,13/ ("coupled loop form") dienen, deren $\underline{A}$ -Matrix von der Form

$$\underline{A} = \rho \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \quad (2-13)$$

ist. Das System realisiert ein konjugiert komplexes Polstellenpaar $\lambda_{1,2} = \rho \cos\psi \pm j\rho \sin\psi = \rho e^{\pm j\psi}$. Gemäß $\underline{x}(k+1) = \underline{A} \underline{x}(k)$ geht $\underline{x}(k+1)$ nach (2-13) durch Drehung um den Winkel ψ und Multiplikation mit ρ aus $\underline{x}(k)$ hervor. Für stabile Pollagen mit $|\rho| < 1$ erfüllt die Struktur die Bedingung (2-11), wenn zur Definition

der ZV-Vektor-Norm (2-6) die Diagonalmatrix $\underline{G}$ gleich der Einheitsmatrix gewählt wird. Die Ellipsen in Bild 2.4 gehen dann in Kreise über, deren Radien in jedem Filterschritt um den Faktor ρ abnehmen.

Als Gegenbeispiel soll ein Filterblock zweiten Grades in 2. kanonischer Form dienen. Im Fall konjugiert komplexer Polstellen $\lambda_{1,2} = \rho \cos \psi \pm j \rho \sin \psi$ ist

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -\rho^2 & 2\rho \cos \psi \end{bmatrix} .$$

Bei verschwindendem Eingangssignal und idealer Arithmetik errechnet man leicht die Zustandsvariablen zu

$$\begin{aligned} x_1(k) &= \hat{x} \rho^k \cos [k\psi + \psi_0] \\ x_2(k) &= \hat{x} \rho^{k+1} \cos [(k+1)\psi + \psi_0] , \end{aligned}$$

wobei die Konstanten $\hat{x}$ und ψ_0 durch einen gegebenen Anfangszustand $\underline{x}(0) = [\hat{x} \cos \psi_0, \hat{x} \rho \cos(\psi + \psi_0)]^T$ festgelegt sind. Damit errechnet man den Zusammenhang zwischen x_1 und x_2 als

$$\rho^2 x_1^2 - 2\rho \cos \psi \cdot x_1 x_2 + x_2^2 = \hat{x}^2 \rho^2 \sin^2 \psi .$$

Die Spitze des ZV-Vektors $\underline{x}$ liegt also für festes k auf einer konzentrischen Ellipse, deren Lage und Form durch λ und ψ , d.h. durch die Polstellen des Filters bestimmt ist. Ferner gilt, daß sich die Halbachsen der Ellipse in jedem Filterschritt um den Faktor ρ ändern. Für den Fall $\rho = 0,9$, $\psi = \pi/4$ ist dies in Bild 2.5 dargestellt. Der wesentliche Unterschied gegenüber Bild 2.4 besteht darin, daß die Ellipsen hier nicht achsensymmetrisch sind. Dem ZV-Vektor läßt sich offenbar auch keine achsensymmetrische Ellipse zuordnen, die mit jedem Filterschritt kleiner wird. Unterstellt man eine Überlaufcharakteristik entsprechend der Zweierkomplement-Kennlinie /6/, so liegt im Beispiel der quantisierte Vektor $\underline{x}(k+1) = [\underline{A} \underline{x}(k)]_Q$ auf einer größeren Ellipse als $\underline{x}(k)$, so daß ein erneuter Überlauf wahrscheinlich ist. Außerdem

wären selbst im idealen linearen Fall mehrere Filterschritte erforderlich, bis der ZV-Vektor wieder im Innern der ursprünglichen Ellipse liegen würde.

Das Auftreten von Überlaufgrenzyklen ist somit nicht auszuschließen.

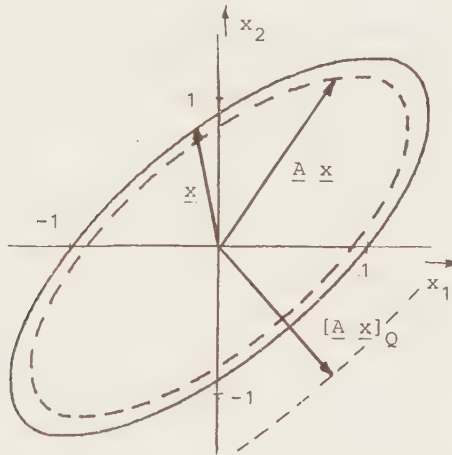


Bild 2.5: Beispiel zum Verhalten des ZV-Vektors bei einem Filterschritt mit Überlauf (2. kan. Form)

Die allgemeine Stabilitätsbedingung (2-11) läßt sich unter Berücksichtigung der Norm-Definition (2-6) noch anders formulieren:

$$\begin{aligned} \underline{x}^T \underline{A}^T \underline{G} \underline{A} \underline{x} &= ||\underline{A} \underline{x}||_{\underline{G}}^2 < ||\underline{x}||_{\underline{G}}^2 = \underline{x}^T \underline{G} \underline{x} \quad \forall \underline{x} \neq 0 \\ \rightarrow \underline{x}^T [\underline{G} - \underline{A}^T \underline{G} \underline{A}] \underline{x} &> 0. \end{aligned} \quad (2-14)$$

Wenn es also gelingt, eine Diagonalmatrix $\underline{G}$ mit positiven Elementen derart zu finden, daß die mit

$$\underline{P} = \underline{G} - \underline{A}^T \underline{G} \underline{A} \quad (2-15)$$

definierte symmetrische Matrix positiv definit ist, so ist das Filter asymptotisch stabil.

Diese Bedingung wurde erstmals in /3/ unter der Annahme idealer

Arithmetik mit $\tilde{x}_v = [\tilde{x}_v]_Q$ im Intervall $-1 < \tilde{x}_v < 1$ angegeben. Turner und Bruton wiesen für den Spezialfall $\underline{G} = \underline{E}$ in /4/ auf die Möglichkeit des Betragsabschneidens zur zusätzlichen Unterdrückung granularer Grenzyklen hin.

Wesentlich ist, daß obige Stabilitätseigenschaft unabhängig von der Aussteuerung bzw. der Skalierung der Zustandsvariablen ist. Dies soll wiederum am Beispiel eines Systems zweiten Grades erläutert werden: Wird anstelle des ZV-Vektors

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

ein durch Skalierung modifizierter Vektor

$$\underline{x}' = \begin{pmatrix} T_1 x_1 \\ T_2 x_2 \end{pmatrix} = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

betrachtet, so ändern die in Bild 2.4 gezeichneten Ellipsen lediglich ihr Halbachsen-Verhältnis. Allgemein ergibt sich mit

$$\underline{x}' = \underline{T}^{-1} \underline{x} ; \quad \underline{T}^{-1} = \text{Diag}\{T_1, T_2, \dots, T_n\}$$

$\underline{T}$ nichtsingulär

$$T_v \neq 0, \infty$$

wegen (2-4) die geänderte $\underline{A}$ -Matrix

$$\underline{A}' = \underline{T}^{-1} \underline{A} \underline{T} \quad \text{bzw.} \quad \underline{A} = \underline{T} \underline{A}' \underline{T}^{-1}$$

Einsetzen in (2-15) liefert

$$\underline{P} = \underline{G} - \underline{T}^{-1} \underline{A}'^T \underline{T} \underline{G} \underline{T} \underline{A}' \underline{T}^{-1}$$

bzw.

$$\underbrace{\underline{T} \underline{P} \underline{T}}_{:= \underline{P}'} = \underbrace{\underline{T} \underline{G} \underline{T}}_{:= \underline{G}'} - \underbrace{\underline{A}'^T \underline{T} \underline{G} \underline{T} \underline{A}'}_{\underline{G}'}$$

Offensichtlich ist $\underline{G}'$ wieder eine Diagonalmatrix mit positiven Elementen und $\underline{P}'$ positiv definit, falls $\underline{P}$ positiv definit ist.

Die in (2-15) formulierte allgemeine Bedingung vereinfacht sich wesentlich für Filter zweiten Grades mit einer $\underline{A}$ -Matrix der Form

$$\underline{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

deren Eigenwerte betragsmäßig kleiner als 1 sind. In diesem Fall gilt nach Mills, Mullis und Roberts /3/ (bzgl. zusätzlicher Erläuterungen siehe Anhang A1):

Eine Diagonalmatrix $\underline{G}$ mit positiv reellen Elementen, für die $\underline{P} = \underline{G} - \underline{A}^T \underline{G} \underline{A}$ positiv definit ist bzw. $||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}}$ gilt, existiert genau dann, wenn entweder

$$A_{12}A_{21} \geq 0 \quad (2-16a)$$

$$\text{oder } A_{12}A_{21} < 0 \quad \text{und} \quad |A_{11} - A_{22}| + \det\{\underline{A}\} < 1 \quad (2-16b)$$

gilt. Die Stabilität des Filters im linearen Fall wird hierbei als gegeben vorausgesetzt.

Anhand dieser Bedingungen sollen zwei Strukturen exemplarisch untersucht werden.

Wir betrachten zunächst ein Normalform-Filter mit der in (2-13) gegebenen $\underline{A}$ -Matrix

$$\underline{A} = \rho \cdot \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}; \quad \lambda_{1,2} = \rho e^{\pm j\psi}$$

Für $\sin\psi \neq 0$ gilt $A_{12}A_{21} < 0$ und $|A_{11} - A_{22}| + \det\{\underline{A}\} = \rho^2 [\cos^2\psi + \sin^2\psi] = \rho^2$. Mit $\rho < 1$ (Stabilität im linearen Fall) ist die Bedingung (2-16b) stets erfüllt.

Für $\sin\psi = 0$ ist $A_{12}A_{21} = 0$ und erfüllt somit (2-16a).

Als weiteres Beispiel möge ein Block zweiten Grades in zweiter kanonischer Form mit

$$\underline{A} = \begin{bmatrix} 0 & 1 \\ -\rho^2 & 2\rho\cos\psi \end{bmatrix} ; \lambda_{1,2} = \rho e^{\pm j\psi}$$

dienen. Für $\rho \neq 0$ gilt $A_{12}A_{21} < 0$ und $|A_{11}-A_{22}| + \det\{A\} = |2\rho\cos\psi| + \rho^2$.

Die Stabilitätsbedingung (2-16b) ist nur für $|2\rho\cos\psi| + \rho^2 < 1$ erfüllt. Dies steht im Einklang mit dem in /6/ angegebenen Ergebnis.

A n m e r k u n g :

Bei allen vorausgegangenen Betrachtungen wurden Quantisierer angenommen, die im Intervall $-1 < \tilde{x}_v < +1$ die Wortlängenverkürzung mittels Betragsabschneiden durchführen. Im Interesse eines geringeren Quantisierungsrauschens ist jedoch eine Rundung der Zustandsvariablen häufig vorzuziehen /5/. In diesem Fall ist die Quantisierereigenschaft $|\underline{x}]_Q| \leq |\tilde{x}_v|$ im Intervall $-1 < \tilde{x}_v < +1$ durch die Beziehung $|\underline{x}]_Q| \leq |\tilde{x}_v| + \frac{Q}{2}$ (Q = Quantisierungsintervall) zu ersetzen. Die Bedingung (2-11) mit $||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}}$ impliziert dann nicht mehr generell die Beziehung (2-9) mit $||\underline{A} \underline{x}]_Q||_{\underline{G}} < ||\underline{x}||_{\underline{G}}$. Sie wird insbesondere dann verletzt, wenn $||\underline{x}||_{\underline{G}}$ klein ist. In praktischen Fällen bleibt die Freiheit von Überlauf-Grenzzyklen erhalten, während sich granulare Grenzzyklen nicht grundsätzlich ausschließen lassen (siehe auch Abschnitt 5.4.2).

2.2.3 Antwortstabilität

Antwortstabilität (forced response stability) bedeutet, daß eine durch Überlauf der Zustandsvariablen hervorgerufene Störung stets abklingt, sofern unter idealen linearen Bedingungen ZV-Werte

$|x_v| > 1$ nur vorübergehend auftreten. Die Freiheit von Überlaufgrenzzyklen ist im allgemeinen nicht hinreichend, um Antwortstabilität sicherzustellen. Bekanntlich können zwar Überlaufgrenzzyklen auch bei Filterblöcken zweiten Grades in einfacher kanonischer Struktur durch Anwendung spezieller Überlauf-Charakteristiken wie der Sättigungs- oder der Dreieckskennlinie vermieden werden /12/; Antwortstabilität ist dagegen, wie Claasen und Kristiansson /7/ gezeigt haben, auch mit solchen Kennlinien nur für einen eingeschränkten Bereich der Polstellenlagen gegeben.

Allgemeine Zusammenhänge zwischen der Freiheit von Überlaufgrenzzyklen und der Antwortstabilität wurden erstmals von Claasen, Mecklenbräuker und Peek in /8/ gefunden. Daran anknüpfend soll im folgenden die Antwortstabilität von Zustandsraumstrukturen gezeigt werden, welche die für Grenzyklusfreiheit angegebene Bedingung (2-11) erfüllen.

Wir unterstellen zunächst einen Eingangsfolgen-Vektor $\underline{v}(k)$, der unter linearen Bedingungen, d.h. idealer Arithmetik, nur vorübergehend Zustandsvariable $|x_{\ell v}(k)| > 1$, $k \leq k_0$, $v = 1(1)n$ (Index ℓ für linearen Fall) hervorruft, während $|x_{\ell v}(k)| < 1$ für alle $k > k_0$ gelten möge. Dies impliziert jedoch nicht, daß im realen nichtlinearen Fall für $k > k_0$ ebenfalls ein überlauffreier Betrieb gewährleistet ist.

Definition:

Ein Filter heißt antwortstabil, wenn die Abweichung des tatsächlich berechneten ZV-Vektors $\underline{x}(k)$ vom idealen Vektor $\underline{x}_{\ell}(k)$ für $k \rightarrow \infty$ gegen Null konvergiert. Dann muß z.B. mit der in (2-6) definierten Norm gelten

$$\lim_{k \rightarrow \infty} \|\underline{x}(k) - \underline{x}_{\ell}(k)\|_{\underline{G}} = 0 \quad (2-17)$$

Für die folgenden Überlegungen sollen nur die durch Überläufe verursachten Quantisierungsfehler berücksichtigt werden, d.h. die

Feinstruktur der Quantisierererkennlinie für $|x_v| < 1$ bleibt unbeachtet. Dies ist sicher zulässig, da Rundungsfehler vergleichsweise geringe Abweichungen bewirken.

Infolge der vorangegangenen Überläufe wird sich für $k > k_0$ der ideal berechnete ZV-Vektor $\underline{x}_\ell(k)$ vom realen Vektor $\underline{x}(k)$ unterscheiden. Nach einem Filterschritt gilt dann mit der Zustandsraum-Beschreibung (2-1)

$$\begin{aligned}\underline{x}(k+1) - \underline{x}_\ell(k+1) &= \underbrace{[\underline{A} \underline{x}(k) + \underline{B} \underline{v}(k)]_Q}_{:= \tilde{\underline{x}}(k+1)} - \underline{A} \underline{x}_\ell(k) - \underline{B} \underline{v}(k) \\ &= |\tilde{\underline{x}}(k+1)|_Q - \underline{x}_\ell(k+1)\end{aligned}\quad (2-18)$$

Wir wollen zunächst die Quantisierung außer acht lassen und erst in einem weiteren Schritt einen möglichen Überlauffehler in die Analyse mit einbeziehen.

Es gilt gemäß (2-18)

$$\begin{aligned}\tilde{\underline{x}}(k+1) - \underline{x}_\ell(k+1) &= \underline{A} \underline{x}(k) + \underline{B} \underline{v}(k) - \underline{A} \underline{x}_\ell(k) - \underline{B} \underline{v}(k) \\ &= \underline{A} [\underline{x}(k) - \underline{x}_\ell(k)]\end{aligned}\quad (2-19)$$

Erfüllt das Filter die Bedingung (2-11): $||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \neq 0$, so folgt aus (2-19)

$$||\tilde{\underline{x}}(k+1) - \underline{x}_\ell(k+1)||_{\underline{G}} < ||\underline{x}(k) - \underline{x}_\ell(k)||_{\underline{G}}, \quad k > k_0 \quad (2-20)$$

Falls es gelingt, eine Quantisierungskennlinie derart zu finden, daß für die Komponenten der Differenzvektoren $(\tilde{\underline{x}} - \underline{x}_\ell)$ und $(\underline{x} - \underline{x}_\ell) = ([\tilde{\underline{x}}]_Q - \underline{x}_\ell)$ die Beziehung

$$|x_v - x_{\ell_v}| = |[\tilde{x}_v]_Q - x_{\ell_v}| \leq |\tilde{x}_v - x_{\ell_v}|, \quad k > k_0 \quad (2-21a)$$

für alle x_{ℓ_v} mit $|x_{\ell_v}| < 1$ erfüllt ist, geht (2-20) über in

$$||\underline{x}(k+1) - \underline{x}_l(k+1)||_{\underline{G}} < ||\underline{x}(k) - \underline{x}_l(k)||_{\underline{G}}, k > k_0. \quad (2-21b)$$

Die in Bild 2.6 dargestellte Kennlinie erfüllt offenbar die Forderung (2-21a). Aus (2-21b) folgt dann unmittelbar die Antwortstabilität des Filters gemäß Definition (2-17).

Wenn also ein Filter die Bedingung (2-11): $||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}}$ erfüllt und gleichzeitig am Eingang der Zustandspeicher die Überlauf-Kennlinie nach Bild 2.6 realisiert wird, so ist dies hinreichend für die Antwortstabilität des Systems.

A n m e r k u n g:

Kennlinien des in Bild 2.6 angegebenen Typs wurden bereits vor mehreren Jahren für Filter zweiten Grades vorgeschlagen /12/. Sie sind hinreichend für die Freiheit von Überlauf-Grenzzyklen (keine Erregung) bei (weitgehend) beliebiger Struktur. Diese Aussage läßt sich jedoch nicht allgemein auf höhergradige Systeme übertragen /53/. Auch folgt aus der Grenzzyklusfreiheit eines Blocks zweiten Grades i.a. nicht dessen Antwortstabilität.

Dagegen sind Strukturen mit der o.a. Eigenschaft (2-11) bereits bei beliebigen Überlaufkennlinien grenzzyklusfrei. Wird nun zusätzlich eine Kennlinie gemäß Bild 2.6 realisiert, so sind diese Filter sogar antwortstabil bei Erregung. Eine Einschränkung bezüglich des Filtergrades besteht hier nicht.

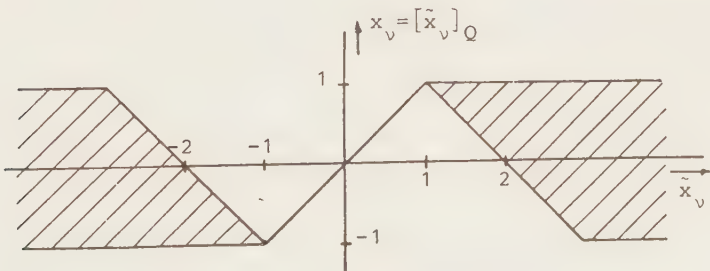


Bild 2.6: Quantisierungskennlinie für antwortstabile Zustandsraum-Strukturen

Bisher wurde lediglich das Verhalten der Zustandsvariablen untersucht. Aus der Antwortstabilität der Zustandsvariablen folgt aber die Antwortstabilität bzgl. des Vektors $\underline{y}(k)$ der Ausgangsfolgen, sofern deren Aussteuerung durch passende Wahl der Matrizen $\underline{C}$ und $\underline{D}$ in (2-1) begrenzt wird. Falls dennoch Zahlenbereichsüberschreitungen auftreten, sind diese unkritisch, da eine Fehlerfortpflanzung wegen der fehlenden Rückkopplung nicht möglich ist.

2.2.4 Ergänzungen

Für Strukturen, bei denen Quantisierer nur am Eingang der Zustandsspeicher auftreten, wurde die Bedingung

$$||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \neq \underline{0} \quad (2-11)$$

als hinreichend für die Antwortstabilität angegeben. Wir wollen nun (2-11) dahingehend erweitern, daß auch das Gleichheitszeichen zulässig ist:¹⁾

$$||\underline{A} \underline{x}||_{\underline{G}} \leq ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n \quad (2-22a)$$

Die Matrix

$$\underline{P} = \underline{G} - \underline{A}^T \underline{G} \underline{A} \quad (2-22b)$$

ist dann positiv (semi)definit (vgl. (2-15)). Im Fall eines Systems zweiten Grades lautet jetzt die Bedingung

$$A_{12} A_{21} \geq 0$$

$$\text{oder } A_{12} A_{21} < 0 \quad \text{und} \quad |A_{11} - A_{22}| + \det\{\underline{A}\} \geq 1 \quad (2-22c)$$

(siehe (2-16) und Anhang A1).

¹⁾ Die Stabilität im linearen Fall sei natürlich vorausgesetzt, so daß das Gleichheitszeichen nur punktuell möglich ist.

Es sind nun die Konsequenzen bzgl. der Überlaufstabilität zu klären.

Wir betrachten den Fall verschwindender Erregung $v(k) \equiv \underline{0}$. Dann gilt bei Stabilität im linearen Fall und idealer Arithmetik natürlich

$$\lim_{k \rightarrow \infty} ||\underline{x}(k)||_{\underline{G}} = 0 . \quad (2-22d)$$

Wir treffen wieder die Annahme, daß die Quantisierung allenfalls zu einer betragsmäßigen Verkleinerung der Komponenten von $\underline{x}$ führen kann (siehe (2-5)). Falls also das ideale Ergebnis $\underline{A} \underline{x}$ nicht mit dem realen $[\underline{A} \underline{x}]_Q$ übereinstimmt, so ist $||[\underline{A} \underline{x}]_Q||_{\underline{G}} < ||\underline{x}||_{\underline{G}}$. (2-22d) gilt folglich auch im realen nichtlinearen Fall.

In der gleichen Weise folgert man aus (2-22a) die Antwortstabilität bei Erregung, wobei die Herleitung entsprechend Abschnitt 2.2.3 geschieht.

2.3 Analyse mit Hilfe eines verallgemeinerten Modells

2.3.1 Modell-Beschreibung

Wir betrachten ein digitales Filter als N-Tor gemäß Bild 2.7a. Das Netzwerk im Innern möge nur lineare arithmetische Operationen wie Additionen, Subtraktionen und Multiplikationen mit konstanten Koeffizienten ausführen. Die externen Tore seien mit den Verzögerungsgliedern und den Quantisierern beschaltet und bilden im übrigen die Ein- und Ausgänge des Systems.

Die Quantisierung werde durch Multiplikation mit einem zeitvariablen von der Eingangsgröße abhängigen Koeffizienten $Q_v(k)$ beschrieben. Faßt man die Eingangsfolgen $a_{1\mu}(k)$, $a_{2\nu}(k)$, $a_{3\lambda}(k)$

in den Vektoren $\underline{a}_1(k)$, $\underline{a}_2(k)$, $\underline{a}_3(k)$ und die Ausgangsfolgen $b_{1\mu}(k)$, $b_{2\nu}(k)$, $b_{3\lambda}(k)$ in den Vektoren $\underline{b}_1(k)$, $\underline{b}_2(k)$, $\underline{b}_3(k)$

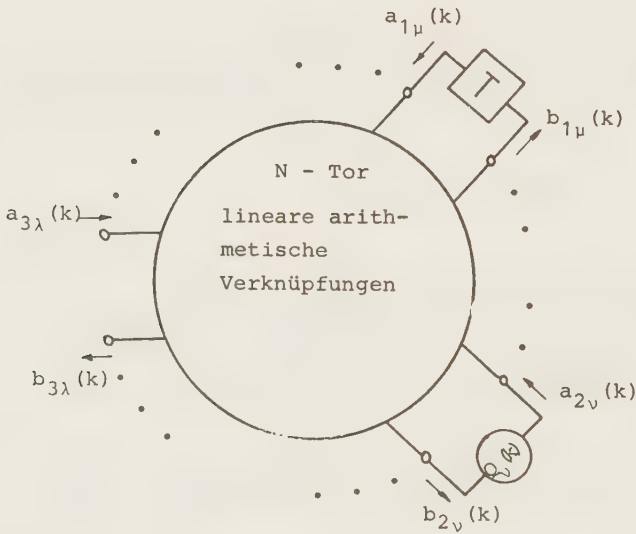


Bild 2.7a: Digitales Filter als N-Tor

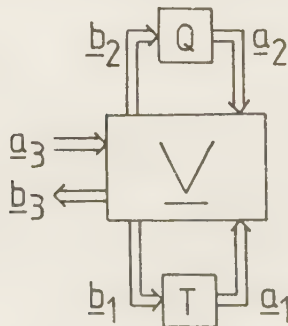


Bild 2.7b: Matrizen-Beschreibung eines allgemeinen Digitalfilters

zusammen, so läßt sich das System folgendermaßen beschreiben (siehe Bild 2.7b):

$$\begin{bmatrix} \underline{b}_1(k) \\ \underline{b}_2(k) \\ \underline{b}_3(k) \end{bmatrix} = \underline{V} \begin{bmatrix} \underline{a}_1(k) \\ \underline{a}_2(k) \\ \underline{a}_3(k) \end{bmatrix} \quad (2-23)$$

$\underbrace{\hspace{1.5cm}}_{:= \underline{b}(k)} \qquad \underbrace{\hspace{1.5cm}}_{:= \underline{a}(k)}$

$\underline{V}$: Matrix mit konstanten reellen Koeffizienten, welche die arithmetischen Operationen des N-Tors beschreibt.

Im Gegensatz zur Zustandsraumstruktur (Abschnitt 2.2) sind hier Quantisierer an beliebiger Stelle zugelassen. Ihre Anordnung ist durch die Verknüpfungsmatrix $\underline{V}$ festgelegt. Die Definition eines Eingangsfolgenvektors $\underline{a}(k)$ und eines Ausgangsfolgenvektors $\underline{b}(k)$ erfolgt auch im Hinblick auf die Streumatrix-Beschreibung der Wellendigitalfilter, die mit dem vorliegenden Modell ebenfalls erfaßt werden.

Die Zustandsraum-Struktur kann als Spezialfall dieser allgemeinen Form betrachtet werden. Die zugehörige Verknüpfungsmatrix $\underline{V}$ wird in Anhang A2, Glgn. (A2-3a) bis (A2-5), hergeleitet und möge hier als Beispiel dienen.

Anhand der in Bild 2.7a dargestellten allgemeinen Struktur wurde bereits von Claasen und Mecklenbräuker ein allgemeiner Zusammenhang zwischen der Antwortstabilität eines Filters und dessen Freiheit von Überlaufgrenzyklen aufgezeigt. Allerdings wurden keine Aussagen über die in (2-23) definierte Verknüpfungsmatrix gemacht.

Demgegenüber sollen hier Bedingungen für $\underline{V}$ formuliert werden, die ein stabiles Überlaufverhalten des Systems gewährleisten.

2.3.2 Grenzzyklusfreiheit

Wir stellen folgende Forderung an die in (2-23) definierte Verknüpfungsmatrix $\underline{V}$:

Es existiere eine konstante Diagonalmatrix $\underline{G}$ mit positiv reellen Elementen, so daß gilt:

$$\underline{b}^T(k) \cdot \underline{G} \cdot \underline{b}(k) \leq \underline{a}^T(k) \cdot \underline{G} \cdot \underline{a}(k) \quad \forall \underline{a}(k) \in \mathbb{R}^N \quad (2-24a)$$

$$\text{bzw.} \quad \underline{a}^T(k) \cdot \underline{V}^T \cdot \underline{G} \cdot \underline{V} \cdot \underline{a}(k) \leq \underline{a}^T(k) \cdot \underline{G} \cdot \underline{a}(k)$$

Mit der Norm-Definition nach Gleichung (2-6) ist (2-24a) äquivalent zu

$$||\underline{b}(k)||_{\underline{G}} \leq ||\underline{a}(k)||_{\underline{G}} \quad \text{bzw.} \quad ||\underline{V} \underline{a}(k)||_{\underline{G}} \leq ||\underline{a}(k)||_{\underline{G}}. \quad (2-24b)$$

Außerdem mögen, wie bereits in Abschnitt 2.2.2, Ungleichung (2-5) gefordert, die Quantisierer die Eigenschaft

$$|a_{2_v}(k)| \leq |b_{2_v}(k)| \quad \text{bzw.} \quad |Q_v(k)| \leq 1 \quad (2-25)$$

besitzen.

Es soll nun gezeigt werden, daß (2-24) zusammen mit (2-25) eine hinreichende Bedingung für Grenzzyklusfreiheit darstellt.

Im Interesse einer übersichtlicheren Schreibweise wird der Laufindex k vorübergehend weggelassen. Wir zerlegen zunächst die Matrix $\underline{G}$ entsprechend der Aufteilung der Vektoren $\underline{a}$ und $\underline{b}$ (Glg. (2-23)) in die Unter-Diagonalmatrizen $\underline{G}_1, \underline{G}_2, \underline{G}_3$, so daß gilt:

$$\begin{aligned} ||\underline{a}||_{\underline{G}}^2 &= \underline{a}_1^T \underline{G}_1 \underline{a}_1 + \underline{a}_2^T \underline{G}_2 \underline{a}_2 + \underline{a}_3^T \underline{G}_3 \underline{a}_3 \\ ||\underline{b}||_{\underline{G}}^2 &= \underline{b}_1^T \underline{G}_1 \underline{b}_1 + \underline{b}_2^T \underline{G}_2 \underline{b}_2 + \underline{b}_3^T \underline{G}_3 \underline{b}_3 \end{aligned} \quad (2-26)$$

Unter Beachtung von (2-24b) erhält man:

$$\underline{b}_1^T \underline{G}_1 \underline{b}_1 + \underline{b}_2^T \underline{G}_2 \underline{b}_2 + \underline{b}_3^T \underline{G}_3 \underline{b}_3 \leq \underline{a}_1^T \underline{G}_1 \underline{a}_1 + \underline{a}_2^T \underline{G}_2 \underline{a}_2 + \underline{a}_3^T \underline{G}_3 \underline{a}_3. \quad (2-27)$$

Aus Bild 2.7b entnimmt man die Beziehung

$$\underline{a}_2 = \underline{Q} \underline{b}_2 \quad \text{mit} \quad \underline{Q} = \text{Diag}[\underline{Q}_1, \underline{Q}_2, \dots] \quad (2-28)$$

Kombiniert man (2-28) mit (2-27) und setzt ferner voraus, daß der Eingangsfolgen-Vektor $\underline{a}_3(k)$ verschwindet, so folgt:

$$\underline{b}_1^T \underline{G}_1 \underline{b}_1 + \underline{b}_2^T \underline{G}_2 \underline{b}_2 + \underline{b}_3^T \underline{G}_3 \underline{b}_3 \leq \underline{a}_1^T \underline{G}_1 \underline{a}_1 + \underline{b}_2^T \underline{Q} \underline{G}_2 \underline{Q} \underline{b}_2. \quad (2-29)$$

Wegen $|\underline{Q}_v| \leq 1$ gemäß (2-25) gilt offenbar

$$\underline{b}_2^T \underline{Q} \underline{G}_2 \underline{Q} \underline{b}_2 \leq \underline{b}_2^T \underline{G}_2 \underline{b}_2,$$

und es folgt mit (2-29):

$$\begin{aligned} \underline{b}_1^T \underline{G}_1 \underline{b}_1 + \underline{b}_2^T \underline{G}_2 \underline{b}_2 + \underline{b}_3^T \underline{G}_3 \underline{b}_3 &\leq \underline{a}_1^T \underline{G}_1 \underline{a}_1 + \underline{b}_2^T \underline{G}_2 \underline{b}_2 \\ \rightarrow \underline{b}_1^T \underline{G}_1 \underline{b}_1 + \underline{b}_3^T \underline{G}_3 \underline{b}_3 &\leq \underline{a}_1^T \underline{G}_1 \underline{a}_1. \end{aligned} \quad (2-30)$$

Hinsichtlich möglicher Grenzzyklen interessiert nur der Fall, daß weder der ZV-Vektor $\underline{a}_1(k) \equiv \underline{0}$ ist, noch der Filterausgangs-Vektor $\underline{b}_3(k)$ identisch verschwindet. Für $\underline{b}_3 \neq \underline{0}$ und $\underline{a}_1 \neq \underline{0}$ folgt aus (2-30):

$$\underline{b}_1^T \underline{G}_1 \underline{b}_1 < \underline{a}_1^T \underline{G}_1 \underline{a}_1 \quad \text{bzw.} \quad ||\underline{b}_1||_{\underline{G}_1}^2 < ||\underline{a}_1||_{\underline{G}_1}^2 \quad (2-31)$$

Wegen $\underline{b}_1(k) = \underline{a}_1(k+1)$ erhält man schließlich

$$||\underline{a}_1(k+1)||_{\underline{G}_1} < ||\underline{a}_1(k)||_{\underline{G}_1} \quad \forall k \quad \text{mit} \quad \underline{b}_3(k) \neq \underline{0} \quad (2-32)$$

Das bedeutet: Eine Zunahme der Norm des ZV-Vektors $\underline{a}_1(k)$ ist ge-

nerell ausgeschlossen. Sie nimmt mit Sicherheit in jedem Filterschritt ab, der einen Ausgangsvektor $\underline{b}_3(k) \neq \underline{0}$ liefert. Am Filterausgang beobachtbare Grenzzyklen sind daher nicht möglich, wenn (2-24) und (2-25) gilt.

Wir suchen nun eine Bedingung für die Verknüpfungsmatrix $\underline{V}$, so daß die Forderung (2-24) erfüllt werden kann. (2-24a) läßt sich schreiben als

$$\underline{a}^T [\underline{G} - \underline{V}^T \underline{G} \underline{V}] \underline{a} \geq 0 \quad \forall \underline{a}.$$

D.h., es muß eine Diagonalmatrix $\underline{G}$ mit positiv reellen Elementen existieren, so daß die mit

$$\underline{P} = \underline{G} - \underline{V}^T \underline{G} \underline{V} \quad (2-33)$$

definierte symmetrische Matrix $\underline{P}$ positiv definit oder positiv semidefinit ist. Es handelt sich um eine ähnliche Bedingung wie sie in Abschnitt 2.2.2 für die $\underline{A}$ -Matrix grenzzuklusfreier Zustandsraum-Strukturen angegeben wurde (siehe Anhang A2).

Es sei noch auf eine wichtige Eigenschaft der oben angegebenen Beschreibung hingewiesen. Sie besteht darin, daß beim Umschalten zweier Teilfilter über ein freies Tor die Bedingung (2-24) erhalten bleibt, sofern diese von den Teilsystemen erfüllt wird. Zum Beweis dieser Aussage spalten wir zunächst von den Vektoren $\underline{a}_3$, $\underline{b}_3$ die Skalare a_{30} , b_{30} ab, welche die Signale am gewählten Verbindungstor beschreiben. Damit lassen sich die beiden Teilfilter folgendermaßen darstellen:

$$\begin{bmatrix} \underline{b}' \\ b_{30} \end{bmatrix} = \underline{V}^{(1)} \cdot \begin{bmatrix} \underline{a}' \\ a_{30} \end{bmatrix}, \quad \begin{bmatrix} \underline{b}'' \\ a_{30} \end{bmatrix} = \underline{V}^{(2)} \cdot \begin{bmatrix} \underline{a}'' \\ b_{30} \end{bmatrix} \quad (2-34)$$

Zur Erläuterung möge Bild 2.7c dienen, wobei die Vektoren $\underline{a}'$, $\underline{b}'$, $\underline{a}''$, $\underline{b}''$ nochmals gemäß Bild 2.7b untergliedert wurden.

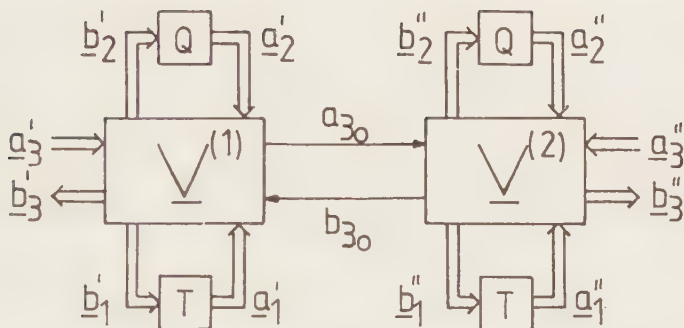


Bild 2.7c: Zusammenschaltung zweier Teilsysteme über ein freies Tor

Die Diagonalmatrizen zur Definition der Vektornorm seien

$$\underline{G}^{(1)} = \begin{bmatrix} \underline{G}' \\ g \end{bmatrix}, \quad \underline{G}^{(2)} = \begin{bmatrix} \underline{G}'' \\ g \end{bmatrix} \quad (2-35)$$

Das skalare Diagonalelement g dient zur Gewichtung der Vektorkomponenten a_{3o} , b_{3o} . g wird bei beiden Teilsystemen als gleich angenommen, was durch eine geeignete Normierung z.B. von $\underline{G}^{(2)}$ immer erreicht werden kann.

Gemäß (2-34) gilt für das neue Gesamtsystem:

$$\begin{bmatrix} \underline{b}' \\ \underline{b}'' \end{bmatrix} = \underline{V}^{(3)} \begin{bmatrix} \underline{a}' \\ \underline{a}'' \end{bmatrix} \quad \text{mit} \quad \underline{V}^{(3)} = \underline{V}^{(3)} \left(\underline{V}^{(1)}, \underline{V}^{(2)} \right). \quad (2-36)$$

Die neue Matrix $\underline{G}^{(3)}$ sei als

$$\underline{G}^{(3)} = \begin{bmatrix} \underline{G}' \\ \underline{G}'' \end{bmatrix} \quad (2-37)$$

definiert. Damit ergibt sich die Vektornorm am Ausgang zu

$$\begin{aligned}
 \left\| \begin{pmatrix} \underline{b}' \\ \underline{b}'' \end{pmatrix} \right\|_{\underline{G}^{(3)}}^2 &= \underline{b}'^T \cdot \underline{G}' \cdot \underline{b}' + \underline{b}''^T \cdot \underline{G}'' \cdot \underline{b}'' \\
 &= \underline{b}'^T \underline{G}' \underline{b}' + b_{3_0} g b_{3_0} - b_{3_0} g b_{3_0} + \underline{b}''^T \underline{G}'' \underline{b}'' + a_{3_0} g a_{3_0} - \\
 &\quad - a_{3_0} g a_{3_0}
 \end{aligned} \tag{2-38}$$

Da nach Voraussetzung

$$\underline{b}'^T \cdot \underline{G}' \cdot \underline{b}' + b_{3_0} g b_{3_0} \leq \underline{a}'^T \cdot \underline{G}' \cdot \underline{a}' + a_{3_0} g a_{3_0}$$

und

$$\underline{b}''^T \cdot \underline{G}'' \cdot \underline{b}'' + a_{3_0} g a_{3_0} \leq \underline{a}''^T \cdot \underline{G}'' \cdot \underline{a}'' + b_{3_0} g b_{3_0}$$

gilt, folgt mit (2-35)

$$\begin{aligned}
 \left\| \begin{pmatrix} \underline{b}' \\ \underline{b}'' \end{pmatrix} \right\|_{\underline{G}^{(3)}}^2 &\leq \underline{a}'^T \underline{G}' \underline{a}' + a_{3_0}^2 g - b_{3_0}^2 g + \underline{a}''^T \underline{G}'' \underline{a}'' + \\
 &\quad + b_{3_0}^2 g - a_{3_0}^2 g \\
 &\leq \underline{a}'^T \underline{G}' \underline{a}' + \underline{a}''^T \underline{G}'' \underline{a}'' = \left\| \begin{pmatrix} \underline{a}' \\ \underline{a}'' \end{pmatrix} \right\|_{\underline{G}^{(3)}}^2
 \end{aligned} \tag{2-39a}$$

bzw.

$$\left\| \underline{b}^{(3)} \right\|_{\underline{G}^{(3)}} \leq \left\| \underline{a}^{(3)} \right\|_{\underline{G}^{(3)}} \tag{2-39b}$$

$$\underline{b}^{(3)} = \begin{bmatrix} \underline{b}' \\ \underline{b}'' \end{bmatrix}, \quad \underline{a}^{(3)} = \begin{bmatrix} \underline{a}' \\ \underline{a}'' \end{bmatrix}.$$

Damit ist der Beweis für obige Behauptung erbracht.

2.3.3 Antwortstabilität

Es soll gezeigt werden, daß Filterstrukturen, welche die Bedingung (2-24) erfüllen und geeignete einfache Quantisierungskennlinien aufweisen, nicht nur antwortstabil im Sinne der in Abschnitt 2.2.3 angegebenen Definition sind, sondern Eigenschaften besitzen, welche in ihrer Bedeutung noch darüber hinaus gehen. Zu diesem Zwecke betrachten wir zwei identische reale Systeme mit unterschiedlichen Anfangszuständen $\underline{a}_1^{(1)}(c)$, $\underline{a}_1^{(2)}(0)$ und unterschiedlicher Erregung $\underline{a}_3^{(1)}(k)$ und $\underline{a}_3^{(2)}(k)$. Für einen Filterschritt gilt dann gemäß (2-23):

$$\underline{b}^{(1)} - \underline{b}^{(2)}(k) = \underline{v} \left[\underline{a}^{(1)}(k) - \underline{a}^{(2)}(k) \right]$$

bzw. (2-40)

$$\begin{bmatrix} \underline{b}_1^{(1)}(k) - \underline{b}_1^{(2)}(k) \\ \underline{b}_2^{(1)}(k) - \underline{b}_2^{(2)}(k) \\ \underline{b}_3^{(1)}(k) - \underline{b}_3^{(2)}(k) \end{bmatrix} = \underline{v} \begin{bmatrix} \underline{a}_1^{(1)}(k) - \underline{a}_1^{(2)}(k) \\ \underline{a}_2^{(1)}(k) - \underline{a}_2^{(2)}(k) \\ \underline{a}_3^{(1)}(k) - \underline{a}_3^{(2)}(k) \end{bmatrix}.$$

Mit der in Gleichung (2-26) gegebenen Darstellung der Vektornorm und bei Beachtung der Bedingung (2-24) erhält man die Beziehung

$$\begin{aligned} & \left| \left| \underline{\Delta b}_1(k) \right| \right|_{\underline{G}_1}^2 + \left| \left| \underline{\Delta b}_2(k) \right| \right|_{\underline{G}_2}^2 + \left| \left| \underline{\Delta b}_3(k) \right| \right|_{\underline{G}_3}^2 \leq \\ & \leq \left| \left| \underline{\Delta a}_1(k) \right| \right|_{\underline{G}_1}^2 + \left| \left| \underline{\Delta a}_2(k) \right| \right|_{\underline{G}_2}^2 + \left| \left| \underline{\Delta a}_3(k) \right| \right|_{\underline{G}_3}^2 \end{aligned} \quad (2-41)$$

mit

$$\begin{aligned} \left| \left| \underline{\Delta b}_v(k) \right| \right|_{\underline{G}_v}^2 &= \left[\underline{b}_v^{(1)}(k) - \underline{b}_v^{(2)}(k) \right]^T \underline{G}_v \left[\underline{b}_v^{(1)}(k) - \underline{b}_v^{(2)}(k) \right] \\ \left| \left| \underline{\Delta a}_v(k) \right| \right|_{\underline{G}_v}^2 &= \left[\underline{a}_v^{(1)}(k) - \underline{a}_v^{(2)}(k) \right]^T \underline{G}_v \left[\underline{a}_v^{(1)}(k) - \underline{a}_v^{(2)}(k) \right] \end{aligned}$$

$$v = 1, 2, 3$$

Wegen $\underline{\Delta b}_1(k) = \underline{\Delta a}_1(k+1)$ folgt

$$\begin{aligned} ||\underline{\Delta a}_1(k+1)||_{\underline{G}_1}^2 &\leq ||\underline{\Delta a}_1(k)||_{\underline{G}_1}^2 + ||\underline{\Delta a}_2(k)||_{\underline{G}_2}^2 - ||\underline{\Delta b}_2(k)||_{\underline{G}_2}^2 + \\ &+ ||\underline{\Delta a}_3(k)||_{\underline{G}_3}^2 - ||\underline{\Delta b}_3(k)||_{\underline{G}_3}^2 . \end{aligned} \quad (2-42)$$

Mit den Anfangszuständen $\underline{a}_1^{(1)}(0)$ und $\underline{a}_1^{(2)}(0)$ folgt

$$\begin{aligned} ||\underline{\Delta a}_1(1)||_{\underline{G}_1}^2 &\leq ||\underline{\Delta a}_1(0)||_{\underline{G}_1}^2 + ||\underline{\Delta a}_2(0)||_{\underline{G}_2}^2 - ||\underline{\Delta b}_2(0)||_{\underline{G}_2}^2 \\ &+ ||\underline{\Delta a}_3(0)||_{\underline{G}_3}^2 - ||\underline{\Delta b}_3(0)||_{\underline{G}_3}^2 , \\ ||\underline{\Delta a}_1(2)||_{\underline{G}_1}^2 &\leq ||\underline{\Delta a}_1(1)||_{\underline{G}_1}^2 + ||\underline{\Delta a}_2(1)||_{\underline{G}_2}^2 - ||\underline{\Delta b}_2(1)||_{\underline{G}_2}^2 \\ &+ ||\underline{\Delta a}_3(1)||_{\underline{G}_3}^2 - ||\underline{\Delta b}_3(1)||_{\underline{G}_3}^2 \\ &\leq ||\underline{\Delta a}_1(0)||_{\underline{G}_1}^2 + \sum_{\lambda=0}^1 \left[||\underline{\Delta a}_2(\lambda)||_{\underline{G}_2}^2 - ||\underline{\Delta b}_2(\lambda)||_{\underline{G}_2}^2 \right] \\ &+ \sum_{\lambda=0}^1 \left[||\underline{\Delta a}_3(\lambda)||_{\underline{G}_3}^2 - ||\underline{\Delta b}_3(\lambda)||_{\underline{G}_3}^2 \right] \end{aligned}$$

usw.

Allgemein erhält man offenbar

$$\begin{aligned} ||\underline{\Delta a}_1(k+1)||_{\underline{G}_1}^2 &\leq ||\underline{\Delta a}_1(0)||_{\underline{G}_1}^2 + \sum_{\lambda=0}^k \left[||\underline{\Delta a}_2(\lambda)||_{\underline{G}_2}^2 - ||\underline{\Delta b}_2(\lambda)||_{\underline{G}_2}^2 \right] \\ &+ \sum_{\lambda=0}^k \left[||\underline{\Delta a}_3(\lambda)||_{\underline{G}_3}^2 - ||\underline{\Delta b}_3(\lambda)||_{\underline{G}_3}^2 \right] \end{aligned} \quad (2-43)$$

Gemäß dem allgemeinen Filtermodell (Abschnitt 2.3.1) gilt:

$$\underline{\Delta a}_2 = \underline{a}_2^{(1)} - \underline{a}_2^{(2)} = \left[\underline{b}_2^{(1)} \right]_Q - \left[\underline{b}_2^{(2)} \right]_Q , \quad (2-44)$$

wobei der Index Q für die Quantisierung der Vektorkomponenten steht. Wir fordern nun eine Quantisierungskennlinie derart, daß die Ungleichung

$$|a_{2\mu}^{(1)} - a_{2\mu}^{(2)}| = \left| \left[\underline{b}_{2\mu}^{(1)} \right]_Q - \left[\underline{b}_{2\mu}^{(2)} \right]_Q \right| \leq |b_{2\mu}^{(1)} - b_{2\mu}^{(2)}| \quad (2-45)$$

stets erfüllt ist. Einfach zu realisierende Beispiele hierfür sind die in Bild 2.8 dargestellten Dreiecks- und Sättigungs-
kennlinien. Aus (2-45) folgt dann unmittelbar

$$||\underline{\Delta a}_2||_{\underline{G}_2}^2 \leq ||\underline{\Delta b}_2||_{\underline{G}_2}^2, \quad (2-46)$$

und die Ungleichung (2-43) läßt sich durch die folgende Beziehung ersetzen:

$$||\underline{\Delta a}_1^{(k+1)}||_{\underline{G}_1}^2 \leq ||\underline{\Delta a}_1^{(0)}||_{\underline{G}_1}^2 + \sum_{\lambda=0}^k ||\underline{\Delta a}_3^{(\lambda)}||_{\underline{G}_3}^2 - \sum_{\lambda=0}^k ||\underline{\Delta b}_3^{(\lambda)}||_{\underline{G}_3}^2 \quad (2-47)$$

Dieses wichtige Ergebnis konnte direkt aus der Stabilitätsbedingung (2-24) abgeleitet werden, wobei lediglich die in Bild 2.8 dargestellten Überlauf-Charakteristiken zugrundegelegt wurden. Es sei nochmals ausdrücklich vermerkt, daß die Herleitung für den realen nichtlinearen Fall erfolgte. Insbesondere wurden keine Einschränkungen bezüglich auftretender Überläufe gemacht.

(2-47) läßt sich nun in unterschiedlicher Weise interpretieren:

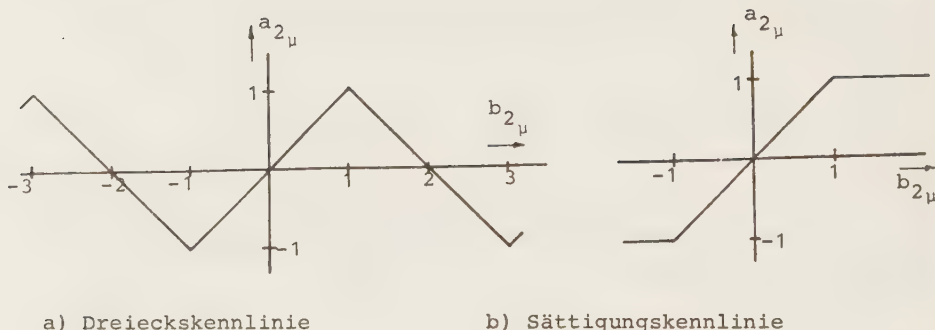


Bild 2.8

Spezialisiert man (2-47) auf den Fall gleicher Erregung, aber unterschiedlicher Anfangszustände, dann folgt mit $\underline{\Delta a}_3(\lambda) \equiv \underline{0}$ $\forall \lambda$:

$$||\underline{\Delta a}_1(k+1)||_{\underline{G}_1}^2 \leq ||\underline{\Delta a}_1(0)||_{\underline{G}_1}^2 - \sum_{\lambda=0}^k ||\underline{\Delta b}_3(\lambda)||_{\underline{G}_3}^2 \quad (2-48)$$

Offenbar verringert sich die Norm des Differenz-ZV-Vektors $\underline{\Delta a}_1(k)$ mit jedem Filterschritt, der einen von Null verschiedenen Differenz-Ausgangsvektor $\underline{\Delta b}_3(k)$ erzeugt. Andererseits muß $\underline{\Delta b}_3(k)$ für $k \rightarrow \infty$ gegen den Nullvektor konvergieren, da die rechte Seite von (2-48) nicht negativ werden kann. Der Einfluß unterschiedlicher Anfangszustände wird also stets abklingen.

Im Zuge einer weiteren Spezialisierung nehmen wir eine Erregung derart an, daß bei geeignetem Anfangszustand $\underline{a}_1^{(2)}(0)$ das Filter 2 überlauffrei arbeitet. Dann folgt aus (2-48) unmittelbar die Überlaufstabilität im Sinne von Abschnitt 2.2.3.

Spezialisiert man (2-47) auf den Fall gleicher Anfangszustände aber unterschiedlicher Erregung, dann folgt mit $\underline{\Delta a}_1(0) = \underline{0}$ und $||\underline{\Delta a}_1(k+1)||_{\underline{G}_1}^2 \geq 0$:

$$\sum_{\lambda=0}^k ||\underline{\Delta b}_3(\lambda)||_{\underline{G}_3}^2 \leq \sum_{\lambda=0}^k ||\underline{\Delta a}_3(\lambda)||_{\underline{G}_3}^2. \quad (2-49)$$

Interpretiert man die Summenausdrücke als Energie der Differenzvektoren am Ein- und Ausgang, dann besagt (2-49), daß die Differenzenergie am Ausgang stets kleiner oder gleich der Differenzenergie am Eingang ist. Insbesondere unterscheiden sich die Ausgangsfolgen beliebig wenig voneinander, wenn die Differenz der Eingangsfolgen hinreichend klein ist. Diese Stetigkeit des Übertragungsverhaltens ist für lineare Systeme selbstverständlich. Sie stellt jedoch im hier betrachteten nichtlinearen Fall eine Besonderheit dar. Hieraus resultiert die folgende wichtige Konsequenz:

Die Fehlerleistung am Ausgang bleibt auch dann klein, wenn die Erregung des Filters permanent zu geringen Übersteuerung führt.

Allerdings ist zu berücksichtigen, daß i.a. die Eingangs- und Ausgangsfolgen an unterschiedlichen Toren auftreten und entsprechend der verwendeten Vektornorm unterschiedlich gewichtet werden.

Die Eigenschaft (2-47) wird von Meerkötter /9/ im Fall der Wellendigitalfilter als "inkrementale Pseudopassivität" bezeichnet und als besonderes Merkmal dieser Strukturen herausgestellt. Wie sich aber noch zeigen wird, erfüllen auch die in Abschnitt 2.2 betrachteten antwortstabilen Zustandsraum-Strukturen mit der

Eigenschaft $||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \neq 0$ die Bedingung (2-24) und können daher ebenfalls als inkremental pseudopassiv bezeichnet werden. Voraussetzung ist hier wie auch im Fall der Wellendigitalfilter eine Überlaufkennlinie wie z.B. in Bild 2.8 angegeben.

Die Untersuchungen im vorliegenden Abschnitt haben nur Quantisierungsfehler infolge Überlauf berücksichtigt, während die Feinquantisierung im Bereich -1 bis +1 vernachlässigt wurde. Diese granularen Fehler machen sich als zusätzliches Rundungsrauschen bemerkbar. Sie ändern grundsätzlich nichts an der Überlaufstabilität des Systems.

2.3.4 Zusammenhang mit den Eigenschaften der Systemmatrix $\underline{A}$

In den vorangegangenen Abschnitten wurden hinreichende Bedingungen für die Überlaufstabilität von Digitalfiltern angegeben. Im ersten Fall wurden Strukturen betrachtet, bei denen die Quantisierer (bzw. Überlaufkennlinien) nur an den Eingängen der Zustandsspeicher auftreten. Hierbei waren die Eigenschaften der Systemmatrix $\underline{A}$ (Zustandsraumbeschreibung) maßgebend und es wurden die Forderungen:

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n \quad (2-22a)$$

oder
$$||\underline{A} \underline{x}||_{\underline{G}_1} < ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \neq 0 \quad (2-11)$$

aufgestellt.

Im Zuge einer Verallgemeinerung wurden Strukturen betrachtet, bei denen nun die Quantisierer an beliebigen Stellen zugelassen waren. Die Beschreibung erfolgte mittels einer Verknüpfungsmatrix $\underline{V}$, für die

$$||\underline{v} \underline{a}||_{\underline{G}} \leq ||\underline{a}||_{\underline{G}} \quad \forall \underline{a} \in \mathbb{R}^N \quad (2-24b)$$

gefordert wurde.

Es soll nun untersucht werden, inwieweit (2-11) und (2-22a) aus (2-24b) folgen bzw. umgekehrt.

Wir setzen zunächst die Gültigkeit von (2-24b) voraus. Gemäß (2-30) erhält man dann für verschwindende Erregung

$$||\underline{b}_1(k)||_{\underline{G}_1}^2 + ||\underline{b}_3(k)||_{\underline{G}_3}^2 \leq ||\underline{a}_1(k)||_{\underline{G}_1}^2 \quad \forall \underline{a}_1(k) \in \mathbb{R}^n. \quad (2-50)$$

Mit der Bezeichnung $\underline{x}(k)$ für den ZV-Vektor gilt

$$\underline{x}(k) = \underline{a}_1(k) \quad , \quad \underline{x}(k+1) = \underline{b}_1(k) \quad .$$

Gemäß (2-50) ergibt sich

$$||\underline{x}(k+1)||_{\underline{G}_1} \leq ||\underline{x}(k)||_{\underline{G}_1} \quad \forall \underline{x}(k) \in \mathbb{R}^n \quad (2-51a)$$

bzw.

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n \quad (2-51b)$$

Das Gleichheitszeichen ist für spezielle $\underline{a}_1(k) = \underline{x}(k) \neq \underline{0}$ möglich, wenn in (2-50) das Gleichheitszeichen gilt und zusätzlich $\underline{b}_3(k) = \underline{0}$ wird.

Der umgekehrte Schluß von (2-11) bzw. (2-51b) auf (2-24b) ist dagegen im allgemeinen nicht möglich. Dies soll an einem einfachen Beispiel zweiten Grades mit

$$\underline{A} = \begin{bmatrix} -\frac{3}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}; \quad \underline{y}(k) = \underline{c}^T \cdot \underline{x}(k) \quad (2-52) \\ \underline{c}^T = [1 \quad 0]$$

gezeigt werden. $\underline{A}$ besitzt die Eigenwerte $\lambda_{1,2} = \pm \frac{1}{\sqrt{2}}$.

Wir überprüfen die Bedingung (2-22a) und erhalten mit $\underline{G} = \underline{E}$

$$||\underline{x}||_{\underline{E}}^2 - ||\underline{A} \underline{x}||_{\underline{E}}^2 = \frac{3}{8} [x_1^2 - 2x_1x_2 + x_2^2] = \frac{3}{8} [x_1 - x_2]^2 \geq 0$$

bzw.

$$||\underline{A} \underline{x}||_{\underline{E}} \leq ||\underline{x}||_{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n.$$

Das Gleichheitszeichen ist für $\underline{x} \neq \underline{0}$ mit $x_1 = x_2$ möglich. Mit dem Anfangszustand $\underline{x}(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ gilt bei verschwindender Erregung

$$\underline{x}(1) = \underline{A} \underline{x}(0) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$||\underline{x}(1)||_{\underline{G}_1} = ||\underline{x}(0)||_{\underline{G}_1}, \quad \underline{G}_1 \text{ beliebig} \quad (2-54)$$

Wegen $y(0) = \underline{c}^T \underline{x}(0) = 1$ gilt ferner

$$||\underline{x}(1)||_{\underline{G}_1}^2 + g \cdot y^2(0) = ||\underline{x}(1)||_{\underline{G}_1}^2 + ||y(0)||_g^2 > ||\underline{x}(0)||_{\underline{G}_1}^2, \quad (2-55)$$

$\underline{G}_1 =$ beliebige positive Diagonalmatrix; $g > 0$, beliebig.

Dies steht aber im Widerspruch zur Beziehung (2-47), welche die inkrementale Pseudopassivität beschreibt. Setzt man nämlich dort

$$\begin{aligned} \underline{\Delta a}_1(k) &= \underline{a}_1^{(1)}(k) = \underline{x}(k), \quad \underline{\Delta a}_3(k) = \underline{a}_3^{(1)}(k) = \begin{bmatrix} v(k) \\ 0 \end{bmatrix} = \underline{0}, \\ \underline{\Delta b}_3(k) &= \underline{b}_3^{(1)}(k) = \begin{bmatrix} 0 \\ y(k) \end{bmatrix}, \end{aligned}$$

so folgt für $k = 0$

$$||\underline{x}(1)||_{\underline{G}_1}^2 + ||\begin{pmatrix} 0 \\ y(0) \end{pmatrix}||_{\underline{G}_3}^2 \leq ||\underline{x}(0)||_{\underline{G}_1}^2.$$

Mit $\underline{G}_3 = \text{Diag}\{g, g\}$ erhält man

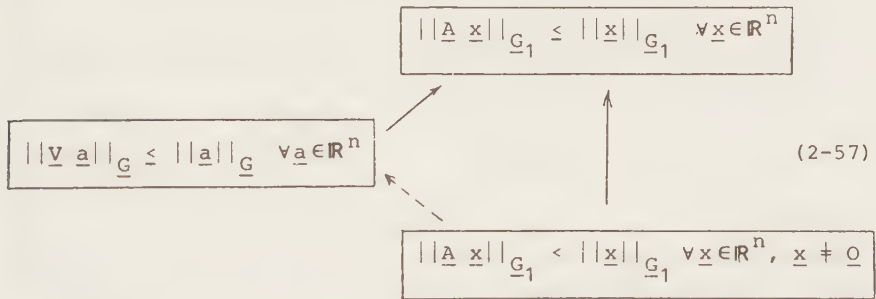
$$||\underline{x}(1)||_{\underline{G}_1}^2 + ||\underline{y}(0)||_g^2 \leq ||\underline{x}(0)||_{\underline{G}_1}^2. \quad (2-56)$$

Da (2-47) und somit (2-56) aus der Bedingung

$||\underline{V} \underline{a}||_{\underline{G}} \leq ||\underline{a}||_{\underline{G}} \quad \forall \underline{a} \in \mathbb{R}^N$ folgt, kann umgekehrt bei Gültigkeit von (2-55) nicht gleichzeitig $||\underline{V} \underline{a}||_{\underline{G}} \leq ||\underline{a}||_{\underline{G}} \quad \forall \underline{a} \in \mathbb{R}^N$ erfüllt sein.

Nehmen wir dagegen an, daß (2-11) gilt, das Gleichheitszeichen also für $\underline{x} \neq 0$ ausgeschlossen ist, so läßt sich zeigen, daß dann immer (2-24b) folgt (Anhang A2,I). Voraussetzung ist hierbei, daß die Quantisierer am Eingang der Zustandsspeicher auftreten.

Die gefundenen Zusammenhänge lassen sich schematisch als



darstellen. Der gestrichelte Pfeil soll auf die genannte Einschränkung bezüglich der Lage der Quantisierer hinweisen.

Weitere Aussagen finden sich im nächsten Unterabschnitt.

2.3.5 Modifikation des allgemeinen Modells

Die Stabilitätsanalyse mit Hilfe des allgemeinen Filtermodells versagt offenbar bei gewissen Zustandsraumstrukturen mit der Eigenschaft $||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1}$, wobei das Gleichheitszeichen für spezielle $\underline{x} \neq \underline{0}$ auftreten kann (siehe Abschnitt 2.3.4).

Diese Schwierigkeit läßt sich aber umgehen, wenn man die Betrachtung auf die Zustandsvariablen und die Eingangssignale beschränkt.

Dazu sind die zum Ausgangssignalvektor $\underline{b}_3(k)$ gehörenden Zeilen von $\underline{V}$ gleich Null zu setzen (siehe Glg. (2-23)). Bezeichnen wir die so entstandene neue Verknüpfungsmatrix mit $\underline{V}_*$, so geht die Stabilitätsbedingung (2-24a,b) in

$$||\underline{V}_* \underline{a}(k)||_{\underline{G}} \leq ||\underline{a}(k)||_{\underline{G}} \quad \forall \underline{a}(k) \in \mathbb{R}^N \quad (2-58)$$

über. Dies entspricht der Forderung, daß eine positive Diagonalmatrix $\underline{G}$ existieren muß, derart, daß mit

$$\underline{P} = \underline{G} - \underline{V}_*^T \underline{G} \underline{V}_* \quad (2-59)$$

eine positiv (semi-) definite Matrix vorliegt (vgl. (2-33)). Hieraus folgert man Grenzzzyklusfreiheit und Antwortstabilität, wobei die Herleitungen denen aus Abschnitt 2.3.2 und 2.3.3 entsprechen. Es ist lediglich $\underline{b}_3(k) \equiv \underline{0}$ zu setzen. Dies bedeutet bzgl. der Grenzzzyklusfreiheit, daß die Ungleichungen (2-30) und (2-32) in

$$\underline{b}_1^T \underline{G}_1 \underline{b}_1 \leq \underline{a}_1^T \underline{G}_1 \underline{a}_1 \quad (2-60)$$

bzw. $||\underline{a}_1(k+1)||_{\underline{G}_1} \leq ||\underline{a}_1(k)||_{\underline{G}_1} \quad \forall \underline{a}_1(k) \in \mathbb{R}^n \quad (2-61)$

übergehen. Dies sichert ein Abklingen des ZV-Vektors $\underline{x}(k) = \underline{a}_1(k)$, abgesehen von dem bedeutungslosen Fall, daß das Gleichheitszeichen für alle $\underline{a}_1(k)$ gilt.

Die Antwortstabilität resultiert aus (2-47) mit $\Delta b_3(\lambda) \equiv 0$. Das System ist dann zwar nicht mehr im strengen Sinn inkremental pseudopassiv, doch sind die Konsequenzen bzgl. des Stabilitätsverhaltens von geringer Bedeutung. Sie beziehen sich nur auf die unbeschalteten Ausgänge.

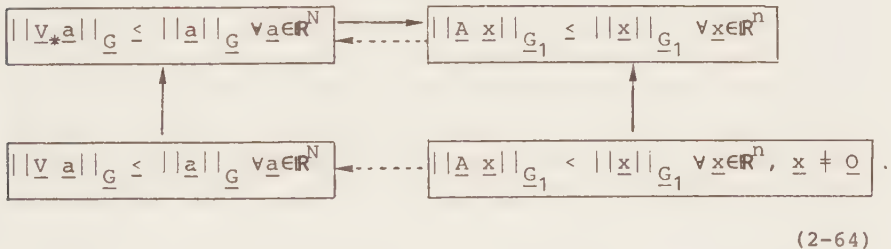
Es soll nun der Zusammenhang mit den Eigenschaften der Systemmatrix $\underline{A}$ aufgezeigt werden.

Aus (2-61) folgt mit $\underline{x}(k) = \underline{a}_1(k)$ und $\underline{x}(k+1) = \underline{A} \underline{x}(k)$:

$$\|\underline{A} \underline{x}\|_{\underline{G}_1} \leq \|\underline{x}\|_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n \quad (2-63)$$

Der umgekehrte Schluß, daß Zustandsraumstrukturen mit der Eigenschaft (2-63) auch die Bedingung (2-58) erfüllen, gilt nun ebenfalls (Anhang A2, II).

Insgesamt ergibt sich mit (2-57) die ergänzte Darstellung der Zusammenhänge



Der gestrichelte Pfeil deutet jeweils an, daß die Umkehrung nur bei Quantisierung am Eingang der Zustandsspeicher möglich ist.

2.4 Zusammenfassung

Es wurden hinreichende Bedingungen für die Antwortstabilität und Grenzzykusfreiheit digitaler Filter formuliert. Dazu wurde in Abschnitt 2.2 zunächst der bekannte Ansatz über die Zustandsraumbeschreibung erläutert. Als wesentliches Ergebnis wurden die Bedingungen (2-11) bzw. (2-15) angegeben, deren Einhaltung bei Quantisierungskennlinien gemäß Bild 2.6 zum gewünschten Stabilitätsverhalten führen. Allerdings gilt das nur, wenn die Quantisierung unmittelbar am Eingang des Zustandsspeichers erfolgt. Aus diesem Grund wurde in Abschnitt 2.3 ein verallgemeinertes Filtermodell zugrundegelegt. Die dazu in (2-24) bzw. (2-33) gefundenen Bedingungen führten bei Verwendung von Quantisierern mit der Dreiecks- oder der Sättigungskennlinie schließlich zur wichtigen Eigenschaft der "inkrementalen Pseudopassivität". Sie schließt die Antwortstabilität und die Grenzzykusfreiheit ein, geht aber in ihrer Bedeutung noch darüber hinaus: So klingt auch bei nicht überlauffreiem Filterbetrieb der Einfluß eines Anfangszustandes stets ab. Ferner ist das Übertragungsverhalten in dem Sinne stetig, daß kleine Änderungen des Eingangssignals nur kleine Änderungen des Ausgangssignals zur Folge haben; d.h. z.B. auch, daß permanent auftretende geringfügige Übersteuerungen nur geringfügige Verzerrungen verursachen können. Die inkrementale Pseudopassivität ist zwar bereits von den Wellendigitalfiltern bekannt, doch konnte nun gezeigt werden, daß auch antwortstabile Zustandsraum-Strukturen mit $||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}}, \underline{x} \in \mathbb{R}^n, \underline{x} \neq \underline{0}$ diese Eigenschaft besitzen.

Bei Verzicht auf die inkrementale Pseudopassivität im strengen Sinn konnten die o.a. Stabilitätsbedingungen etwas weiter gefaßt werden (Abschnitte 2.2.4 und 2.3.5), wobei die Aussagen bezüglich des Stabilitätsverhaltens nur geringfügig zu modifizieren sind.

Es sei vermerkt, daß in der Regel nur Quantisierungsfehler infolge Übersteuerung berücksichtigt wurden, wobei die aus der endlichen Wortbreite der Zahlendarstellung resultierende Feinstruktur der Quantisierungskennlinie vernachlässigt wurde. Dies gilt insbesondere für die Untersuchung der Stabilität des erregten Systems.

Für den Sonderfall verschwindender Erregung können dagegen die Rundungsfehler mit in die Betrachtung einbezogen werden: Systeme mit den oben diskutierten Stabilitätseigenschaften sind nämlich bei Betragsabschneiden zusätzlich frei von granularen Grenzyklen. Diese Aussage gilt jedoch nicht mehr allgemein, wenn wir periodische Eingangssignale zulassen /51/. Darüber hinaus hat das Betragsabschneiden ein erhöhtes Rundungsgeräusch zur Folge.

3. ÜBERLAUFSTABILE FILTERSTRUK- TUREN - BESCHREIBUNG UND VER- GLEICHENDE UNTERSUCHUNGEN

3.1 Vorbemerkungen

Im vorangegangenen Kapitel wurden allgemeine Bedingungen für überlaufstabile Digitalfilter formuliert und daraus resultierende Eigenschaften erläutert. Es sollen nun eine Reihe von bekannten Filterstrukturen angegeben werden, welche den aufgestellten Bedingungen genügen und damit hinsichtlich ihres Übersteuerungsverhaltens in enger Beziehung zueinander stehen. Wir unterscheiden zwischen Zustandsraumfiltern, allgemeinen Wellendigitalfiltern nach Fettweis und Leiterstrukturen.

Die Zielsetzung im vorliegenden Abschnitt läßt sich folgendermaßen charakterisieren:

1. Anhand von Beispielen soll die Anwendung der Stabilitätstheorie (Kap. 2) demonstriert werden. Insbesondere soll auch die Brauchbarkeit des neu eingeführten verallgemeinerten Filtermodells für die Stabilitätsanalyse z.B. von Wellendigitalfiltern aufgezeigt werden.
2. Die Besonderheiten der Systemmatrix $\underline{A}$ sollen für die betrachteten Strukturen erarbeitet und, soweit möglich, Verwandtschaften aufgezeigt werden.

Als Basis für die genannten Betrachtungen wird der jeweilige Filteraufbau erläutert. Insbesondere im Fall der Zustandsraumstrukturen erfolgt eine sehr ausführliche Beschreibung des Entwurfs, welche teilweise über die Darstellung in der Literatur hinausgeht bzw. diese ergänzt.

3.2 Zustandsraum - Strukturen

3.2.1 Aussteuerung und Rundungsrauschen

Wegen des auf das Intervall $[-1, +1]$ beschränkten Zahlenbereiches ist es erforderlich, die Aussteuerung zu begrenzen, so daß Überläufe vermieden werden oder zumindest ihre Auftretswahrscheinlichkeit klein ist. Bei einer Darstellung der Zahlen im Zweierkomplement, die hier stets angenommen wird, kann die Überprüfung des Dynamikbereichs auf die Signale am Eingang der Multiplizierer beschränkt werden /10/. Diese kritischen Stellen sind im Fall der Zustandsraum-Strukturen offenbar die Zustandsspeicher.

Andererseits treten i.a. stochastische Fehler infolge Rundung der Zustandsvariablen auf. Diese erzeugen am Filterausgang ein Störgeräusch, dessen relative Leistung umso größer ist, je weiter die Aussteuerung der Zustandsvariablen verringert wird. Das Ziel beim Entwurf der nachfolgend beschriebenen Strukturen ist daher u.a., bei gegebener Aussteuerung im Innern ein minimales Rundungsrauschen zu erreichen.

Eine Zusammenstellung der theoretischen Grundlagen zur Behandlung der genannten Probleme findet sich im Anhang A3.

3.2.2 Die Normalform - Struktur

Diese Struktur wurde von Barnes und Fam /13/ vorgeschlagen. Hierbei war das primäre Ziel, eine Filter-Realisierung zu finden, welche bei normaler Zweierkomplement-Kennlinie frei von Überlauf-Grenzyklen ist. Barnes /14/ zeigte darüber hinaus, wie für derartige Strukturen das Rundungsgeräusch minimiert werden kann. Weitere Ergebnisse sind in /15/ - /19/ publiziert.

Das primäre Kennzeichen der Struktur ist die besondere Form der A-Matrix. Es wird gefordert, daß A reell normal sein soll und somit die folgenden Eigenschaften besitzt /20/:

$$\underline{A}^T \underline{A} = \underline{A} \underline{A}^T \quad (3-1)$$

$$\underline{A} = \underline{U} \underline{\Lambda} \underline{U}^{*T} \quad (3-2)$$

$$\underline{\Lambda} = \text{Diag}\{\lambda_1, \dots, \lambda_n\}$$

= Diagonalmatrix der Eigenwerte von $\underline{A}$

$$\underline{U} = \text{unitäre Modalmatrix mit } \underline{U}^{*T} \underline{U} = \underline{E}$$

$$\underline{A}^* = \underline{A} \quad (3-3)$$

Wir betrachten den Rayleigh Quotienten

$$R[\underline{x}] = \frac{\underline{x}^T \underline{A}^T \underline{A} \underline{x}}{\underline{x}^T \underline{x}} = \frac{||\underline{A} \underline{x}||_{\underline{E}}^2}{||\underline{x}||_{\underline{E}}^2}; \quad \underline{x} \in \mathbb{R}^n \quad (3-4)$$

$||\dots||_{\underline{E}}$ bezeichnet hier die euklidische Vektornorm. Sie geht aus der in (2-6) definierten Norm hervor, wenn dort die Diagonalmatrix $\underline{G}$ gleich der Einheitsmatrix gesetzt wird.

Nach /20/ sind die Extremwerte von $R[\underline{x}]$ gleich den positiv reellen Eigenwerten κ_v^2 , $v = 1(1)n$, der Matrix $\underline{A}^T \underline{A}$, wobei $\underline{x}_v$ der jeweils zugehörige reelle Eigenvektor ist. Für normale Matrizen $\underline{A}$ folgt aus (3-2) und (3-3)

$$\begin{aligned} \underline{A}^T \underline{A} &= \underline{A}^{*T} \underline{A} = \underline{U} \underline{\Lambda}^* \underline{U}^{*T} \underline{U} \underline{\Lambda} \underline{U}^{*T} = \underline{U} \underline{\Lambda}^* \underline{\Lambda} \underline{U}^{*T} \\ \text{mit } \underline{\Lambda}^* \underline{\Lambda} &= \text{Diag}\{|\lambda_1|^2, \dots, |\lambda_n|^2\} \\ &= \text{Diag}\{\kappa_1^2, \dots, \kappa_n^2\}. \end{aligned} \quad (3-5)$$

Mit (3-4) erhält man

$$||\underline{A} \underline{x}||_{\underline{E}}^2 \leq \max_v \{|\lambda_v|^2\} \cdot ||\underline{x}||_{\underline{E}}^2 \quad \forall \underline{x} \in \mathbb{R}^n \quad (3-6a)$$

bzw.

$$\|\underline{A} \underline{x}\|_{\underline{E}} \leq \max_{\underline{v}} \{|\lambda_{\underline{v}}|\} \cdot \|\underline{x}\|_{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n \quad (3-6b)$$

Ist das System im linearen Fall stabil, so folgt mit $|\lambda_{\underline{v}}| < 1$,
 $\underline{v} = 1(1)n$:

$$\|\underline{A} \underline{x}\|_{\underline{E}} < \|\underline{x}\|_{\underline{E}} \quad \forall \underline{x} \neq \underline{0} \quad (3-7)$$

Das Filter ist somit nach Kap. 2 inkremental pseudopassiv, sofern als Überlaufcharakteristik die Dreiecks- oder Sättigungskennlinie verwendet wird.

Für die zur euklidischen Vektornorm gehörende Schrankennorm (Supremumsnorm) der Matrix $\underline{A}$ gilt hier offenbar

$$\sup[\underline{A}] = \max_{\underline{x}} \frac{\|\underline{A} \underline{x}\|_{\underline{E}}}{\|\underline{x}\|_{\underline{E}}} = \max_{\underline{v}} \{|\lambda_{\underline{v}}|\} \quad (3-8)$$

Andererseits ist für beliebige Matrizen $\underline{A}$ mit den Eigenwerten $\lambda_{\underline{v}}$
 /20/

$$\max_{\underline{v}} \{|\lambda_{\underline{v}}|\} \leq \sup[\underline{A}] \quad (3-9)$$

$\sup[\underline{A}]$ ist demnach für normale Matrizen $\underline{A}$ minimal, wobei in (3-9) das Gleichheitszeichen gilt.

Im folgenden werden zwei Beispiele einer normalen $\underline{A}$ -Matrix angegeben.

Wir betrachten zunächst den Fall reeller Eigenwerte $\lambda_{\underline{v}}$. Gemäß (3-2) gilt dann

$$\underline{A} = \underline{U} \underline{\Lambda} \underline{U}^T \quad (3-10)$$

$\underline{\Lambda}$ = reelle Diagonalmatrix der Eigenwerte

$\underline{U}$ = orthogonale (reell unitäre) Modalmatrix
 mit $\underline{U}^T \underline{U} = \underline{E}$.

Es folgt

$$\underline{A}^T = (\underline{U} \underline{A} \underline{U}^T)^T = \underline{U} \underline{A} \underline{U}^T = \underline{A} . \quad (3-11)$$

$\underline{A}$ ist also symmetrisch. Umgekehrt ist jede symmetrische Matrix $\underline{A}$ wegen $\underline{A}^T \underline{A} = \underline{A} \underline{A}^T$ auch normal (siehe (3-1)). Als zweites Beispiel möge eine (2×2) -Matrix mit konjugiert komplexen Eigenwerten $\lambda_{1,2} = \rho e^{\pm j\psi} = \sigma \pm j\omega$ dienen. Aus der Bedingung $\underline{A} \underline{A}^T = \underline{A}^T \underline{A}$ (3-1) folgt nach einer kurzen Zwischenrechnung

$$\underline{A} = \rho \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix} = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} \quad (3-12)$$

Hier gilt speziell

$$||\underline{A} \underline{x}||_E = \rho ||\underline{x}||_E$$

Bei gegebener Übertragungsfunktion mit einfachen Polstellen läßt sich nun stets eine Zustandsraumstruktur angeben, deren $\underline{A}$ -Matrix normal ist. Darüber hinaus existieren i.a. Freiheitsgrade, die zur Minimierung des Rundungsrauschens genutzt werden können. Eine ausführliche Beschreibung des Strukturentwurfs bei gegebener Übertragungsfunktion findet sich in Anhang A4. Hierbei wird eine Skalierung der Zustandsvariablen nach der L_2 -Norm vorausgesetzt.

Soll zur Skalierung eine andere Norm verwendet werden, so ist dies stets durch eine nachträgliche Transformation des Zustandsvektors mit einer geeignet gewählten reellen Diagonalmatrix $\underline{T}$ möglich, wobei $\underline{x}$ in $\underline{x}' = \underline{T}^{-1} \underline{x}$ und $\underline{A}$ in $\underline{A}' = \underline{T}^{-1} \underline{A} \underline{T}$ übergeht. Im folgenden soll geklärt werden, welchen Einfluß diese Maßnahme auf die Eigenschaften der $\underline{A}$ -Matrix hat. Zunächst gilt für das ursprüngliche System gemäß (3-6)

$$||\underline{A} \underline{x}||_E^2 \leq \max \{ |\lambda_v|^2 \} ||\underline{x}||_E^2$$

bzw.

$$\underline{x}^T \underline{A}^T \underline{A} \underline{x} \leq \max \{ |\lambda_v|^2 \} \underline{x}^T \underline{x} . \quad (3-13)$$

Nach Einsetzen von $\underline{x} = \underline{T} \underline{x}'$ und $\underline{A} = \underline{T} \underline{A}' \underline{T}^{-1}$ folgt

$$\begin{aligned} \underline{x}'^T \underline{T}^T \underline{T}^{-1} \underline{A}'^T \underline{T} \underline{T} \underline{A}' \underline{T}^{-1} \underline{T} \underline{x}' &= \underline{x}'^T \underline{A}'^T \underline{T}^2 \underline{A}' \underline{x}' = (\underline{A}' \underline{x}')^T \underline{T}^2 (\underline{A}' \underline{x}') \\ &\leq \max \{ |\lambda_v|^2 \} \underline{x}'^T \underline{T}^2 \underline{x}' \end{aligned}$$

bzw.

$$\| \underline{A}' \underline{x}' \|_{\underline{G}=\underline{T}^2} \leq \max \{ |\lambda_v|^2 \} \| \underline{x}' \|_{\underline{G}=\underline{T}^2} \quad (3-14)$$

Das transformierte Filter ist also ebenfalls inkremental pseudo-passiv, wenn auch $\underline{A}'$ i.a. nicht mehr normal ist. Die Bedingung einer normalen Matrix $\underline{A}$ ist demnach hinreichend, aber nicht notwendig.

Obiges Ergebnis ist ein Spezialfall der bereits in Kapitel 2 gefundenen Aussage, daß bei Transformation mit einer beliebigen reellen Diagonalmatrix die Überlaufstabilität erhalten bleibt.

3.2.3 Die rauschoptimale Zustandsraum-Struktur

3.2.3.1 Vorbemerkung

1976 veröffentlichten Mullis und Roberts /21/ ein analytisches Verfahren zur Bestimmung der rauschoptimalen Zustandsraum-Struktur mit L_2 skalierten Zustandsvariablen (bzgl. der L_2 -Norm siehe Anhang A3). Unabhängig davon erschien 1977 eine Arbeit von Hwang /22/, die prinzipiell zu den gleichen Ergebnissen führte. Von Jackson, Lindgren und Kim /23/ wurden 1979 geschlossene Formeln zur Berechnung der Koeffizienten eines Blockes zweiten Grades angegeben. Schließlich untersuchten Mullis und Roberts den Einfluß einer Frequenztransformation /25/. Sie stellten fest, daß nach einer Transformation mit einem Allpaß n-ten Grades das Rundungs-

rauschen ihrer Struktur n-mal so groß wird. Insbesondere ist z.B. die Störleistung unabhängig von der Bandbreite eines Tiefpasses (Transformation mit $n=1$).

Obwohl der Struktur-Entwurf ohne Berücksichtigung der Überlaufstabilität erfolgte, konnte nachträglich gezeigt werden, daß das System im Spezialfall zweiten Grades überlaufstabil ist /3/. Der folgende Unterabschnitt wird sich daher nur mit diesem Sonderfall beschäftigen.

3.2.3.2 Filterblock zweiten Grades

Nach Mullis und Roberts /21/ ist ein Filter mit L_2 -skalierten Zustandsvariablen genau dann rauschoptimal, wenn die folgenden Bedingungen erfüllt sind:

1. Es existiert eine reelle Diagonalmatrix $\underline{D}_O$, so daß

$$\underline{D}_O^{-1} \underline{K} \underline{D}_O^{-1} = \underline{D}_O \underline{W} \underline{D}_O \quad (3-15a)$$

und

$$2. K_{ii} W_{ii} = K_{jj} W_{jj} \quad i, j \in \{1, \dots, n\} \quad (3-15b)$$

$\underline{K}$ ist hierbei die Kovarianzmatrix der Impulsantworten vom Filtereingang zu den Zustandsvariablen. $\underline{W}$ ist die Kovarianzmatrix der Impulsantworten von den Rundungsstellen zum Ausgang. K_{ii} und W_{ii} bezeichnen jeweils die Hauptdiagonalelemente. Berücksichtigt man, daß bei L_2 -Skalierung alle K_{ii} , $i = 1(1)n$, gleich sind (siehe Anhang A3), so ist (3-15a,b) gleichbedeutend mit

$$\underline{W} = \xi^2 \underline{K}, \quad (3-16)$$

wobei ξ eine beliebige reelle Konstante ist.

Nach /23/ ist diese Bedingung erfüllt, wenn das Filter die Eigenschaften

$$\underline{A}^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \underline{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (3-17a)$$

und

$$\underline{c} = \xi \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \underline{b} \quad \underline{b}, \underline{c}: \text{Koeffizientenvektoren der Zustandsraum-Beschreibung} \quad (3-17b)$$

besitzt. Zum Beweis errechnen wir $\underline{K}$ und $\underline{W}$ aus den im Anhang A3 (Gln. (A3-13), (A3-31)) gegebenen allgemeinen Beziehungen

$$\underline{K} = \sum_{k=0}^{\infty} \underline{A}^k \underline{b} \underline{b}^T \underline{A}^{kT} \quad (3-18a)$$

$$\underline{W} = \sum_{k=0}^{\infty} \underline{A}^{kT} \underline{c} \underline{c}^T \underline{A}^k \quad (3-18b)$$

Durch Einsetzen von (3-17a,b) gewinnt man

$$\begin{aligned} \underline{W} &= \sum_{k=0}^{\infty} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \underline{A}^k \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \xi \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \underline{b} \xi \underline{b}^T \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \underline{A}^{kT} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &= \sum_{k=0}^{\infty} \xi^2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \underline{A}^k \underline{b} \underline{b}^T \underline{A}^{kT} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \\ &= \xi^2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \underline{K} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{aligned} \quad (3-19)$$

Da $\underline{K}$ symmetrisch ist und wegen der Skalierungsbedingung gleiche Hauptdiagonalelemente aufweist, folgt unmittelbar die gewünschte Eigenschaft (3-16).

Es sei vermerkt, daß die Forderung (3-17a,b) allein noch nicht die richtige Aussteuerung sicherstellen. Allerdings ist eine nachträgliche Skalierung durch Transformation mit einer Diagonalmatrix stets möglich, wobei die Gültigkeit von (3-17a,b) erhalten bleibt.

Es soll nun eine Übertragungsfunktion

$$H(z) = \frac{e_1 z + e_0}{z^2 - 2\sigma z + (\sigma^2 + \omega^2)} \quad (3-20)$$

mit den konjugiert komplexen Polstellen $\sigma \pm j\omega$ realisiert werden. Offenbar lässt sich (3-17a) mit einer normalen A-Matrix der Form

$$\underline{A} = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}, \quad \underline{A}^T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \underline{A} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (3-21)$$

erfüllen. Mit

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \underline{c}^T = [c_1, c_2]$$

errechnet man

$$\begin{aligned} H(z) &= \underline{c}^T [z\underline{E} - \underline{A}]^{-1} \underline{b} \\ &= \frac{[b_1 c_1 + b_2 c_2] z - \sigma [b_1 c_1 + b_2 c_2] + \omega [b_2 c_1 - b_1 c_2]}{z^2 - 2\sigma z + [\sigma^2 + \omega^2]} \end{aligned}$$

Unter Berücksichtigung der zu (3-17b) äquivalenten Beziehung

$$\frac{c_1}{c_2} = \frac{b_2}{b_1} \quad \text{bzw.} \quad b_2 = \frac{c_1}{c_2} b_1 \quad (3-22)$$

erhält man

$$H(z) = \frac{2b_1 c_1 z - 2b_1 c_1 \sigma + b_1 \left[\frac{c_1^2}{c_2^2} - c_2 \right] \omega}{z^2 - 2\sigma z + (\sigma^2 + \omega^2)} \quad (3-23)$$

Ein Koeffizientenvergleich mit (3-20) liefert die folgenden Bestimmungsgleichungen

$$\begin{aligned} 2b_1 c_1 &= e_1 \\ -2b_1 c_1 \sigma + b_1 \left[\frac{c_1^2}{c_2^2} - c_2 \right] \omega &= e_0 \end{aligned} \quad (3-24)$$

Nutzt man den hierin verbleibenden Freiheitsgrad zur Wahl von $c_2 = 1$, so ergibt sich die Lösung

$$b_1 = \frac{1}{2\omega} \left[-(e_1\sigma + e_0) \pm \sqrt{e_1^2(\sigma^2 + \omega^2) + 2e_1e_0\sigma + e_0^2} \right]$$

$$c_1 = \frac{e_1}{2b_1}, \quad c_2 = 1 \quad (3-25)$$

$$b_2 = \frac{c_1}{c_2} b_1 = \frac{e_1}{2}$$

Der bei b_1 auftretende Radikand ist in dem hier betrachteten Fall konjugiert komplexer Polstellen stets positiv, so daß b_1 immer reell ist. Ferner läßt sich durch geeignete Vorzeichenwahl erreichen, daß b_1 ungleich Null wird, so daß zur Berechnung von c_1 ein definierter Ausdruck vorliegt.

Das so bestimmte Filter ist allerdings noch nicht skaliert. Dies erfolgt nun durch Transformation mit der Diagonalmatrix

$$\underline{T} = \text{Diag} \{ \delta ||F_1||_2, \quad \delta ||F_2||_2 \}, \quad (3-26)$$

wobei $||F_1||_2$ und $||F_2||_2$ die in Anhang A3 (Glg. (A3-9)) definierten L_2 -Normen der Zustandsvariablen sind. Als endgültige Lösung erhält man

$$\underline{A}' = \underline{T}^{-1} \underline{A} \underline{T}, \quad \underline{b}' = \underline{T}^{-1} \underline{b}, \quad \underline{c}'^T = \underline{c}^T \underline{T} \quad (3-27)$$

Die Betrachtungen im vorliegenden Abschnitt folgten weitgehend den Ausführungen von Jackson, Lindgren und Kim /23/, jedoch wurde für obige Koeffizientenbestimmung ein anderer Weg beschritten, der in /23/ nur am Rande erwähnt wird. Neben einer einfacheren Herleitung wird hierdurch der Zusammenhang mit den in Abschnitt 3.2.2 behandelten Normalform-Strukturen deutlich. Die $\underline{A}$ -Matrix dieser Filter besitzt nämlich im Fall zweiten Grades die Form (3-21) (vgl. 3-12). Der Unterschied besteht lediglich in der nachträglichen Skalierung der rauschoptimalen Form gemäß (3-26), (3-27), was aber die Überlaufstabilität des Systems nicht beeinträchtigt. Es gilt

$$(\underline{A}'\underline{x}')^T \underline{G} (\underline{A}'\underline{x}') = \rho^2 \underline{x}'^T \underline{G} \underline{x}' \quad (3-28)$$

bzw. $||\underline{A}'\underline{x}'||_{\underline{G}} = \rho ||\underline{x}'||_{\underline{G}} \quad \forall \underline{x}' \neq \underline{0}$

$$\rho = \sqrt{\sigma^2 + \omega^2} = \text{Betrag der Polstelle}$$

Für die Normalform-Struktur war $\underline{G}$ gleich der Einheitsmatrix, während hier

$$\underline{G} = \delta^2 \text{Diag}\{||F_1||_2^2, ||F_2||_2^2\} = \underline{T}^2 \quad (3-29)$$

zu setzen ist.

Nach Kapitel 2 ist das Filter inkremental pseudopassiv, wenn als Überlaufcharakteristik die Dreiecks- oder Sättigungskennlinie verwendet wird.

Es soll nun der Fall reeller Polstellen λ_1, λ_2 betrachtet werden. Nach /23/ ist

$$\underline{A} = \begin{bmatrix} \frac{1}{2} (\lambda_1 + \lambda_2) & \frac{1}{2} (\lambda_1 - \lambda_2) \\ \frac{1}{2} (\lambda_1 - \lambda_2) & \frac{1}{2} (\lambda_1 + \lambda_2) \end{bmatrix} \quad (3-30)$$

zu setzen und erfüllt damit die Forderung (3-17a). Die Vektoren $\underline{b}$ und $\underline{c}^T$ erhält man wieder durch Koeffizientenvergleich unter der Nebenbedingung (3-17b) bzw. (3-22). Die Skalierung erfolgt nachträglich mit Hilfe der in (3-26) gegebenen Diagonalmatrix. Allerdings liefert eine derartige Vorgehensweise nicht immer reelle Koeffizienten für $\underline{b}$ und $\underline{c}^T$. In solchen Fällen ist anstelle von (3-17b) die Bedingung

$$\underline{c} = \xi \underline{b}, \quad \xi \in \mathbb{R} \quad (3-31)$$

anzusetzen /23/, welche zusammen mit $\underline{A}^T = \underline{A}$ ebenfalls die gewünschte Eigenschaft (3-16) sichert. Mit (3-18a,b) gilt nämlich:

$$\underline{W} = \sum_{k=0}^{\infty} \underline{A}^{k^T} \underline{c} \underline{c}^T \underline{A}^k = \sum_{k=0}^{\infty} \xi^2 \underline{A}^k \underline{b} \underline{b}^T \underline{A}^{k^T} = \xi^2 \underline{K}.$$

Die Skalierung darf jetzt aber nicht mehr mit einer Diagonalmatrix durchgeführt werden, da dies i.a. die Bedingungen (3-31) zerstören würde. Stattdessen wird mit einer geeigneten orthogonalen Matrix

$$\underline{U} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

transformiert, welche die Aussteuerung der Zustandsvariablen bzgl. der L_2 -Norm gleich macht (siehe Anhang A3, Glg. (A3-25)). Wegen

$$\underline{A}'^T = \left[\underline{U}^T \underline{A} \underline{U} \right]^T = \underline{U}^T \underline{A}^T \underline{U} = \underline{U}^T \underline{A} \underline{U} = \underline{A}'$$

und

$$\underline{c}' = \underline{U}^T \underline{c} = \xi \underline{U}^T \underline{b} = \xi \underline{b}'$$

bleibt sowohl die Symmetrie der $\underline{A}$ -Matrix als auch die Eigenschaft (3-31) erhalten. Eine zusätzliche Erhöhung oder Verminderung der Aussteuerung ist dann durch Transformation mit einer skalaren Konstanten möglich, was ohne Einfluß auf $\underline{A}$ bleibt. Die mit dieser modifizierten Methode gefundene Struktur ist offenbar mit einem rauschoptimalen Normalform-Filter identisch.

Bezüglich der Überlaufstabilität gelten für reelle Polstellen die gleichen Aussagen wie für konjugiert komplexe. Es ist lediglich (3-28) durch

$$(\underline{A}' \underline{x}')^T \underline{G} (\underline{A}' \underline{x}') \leq \max \{ \lambda_1^2, \lambda_2^2 \} \underline{x}'^T \underline{G} \underline{x}' \quad \forall \underline{x}' \neq \underline{0}$$

$$\text{bzw.} \quad (3-32)$$

$$||\underline{A}' \underline{x}'||_{\underline{G}} \leq \max \{ |\lambda_1|, |\lambda_2| \} ||\underline{x}'||_{\underline{G}} \quad \forall \underline{x} \neq \underline{0}$$

zu ersetzen.

Im Fall der zuletzt beschriebenen modifizierten Methode ist $\underline{A}'$ normal, so daß mit $\underline{G} = \underline{E}$ die Vektornorm $||\dots||_{\underline{G}}$ in die euklidische Norm übergeht.

Für den Strukturentwurf wurde eine Skalierung nach der L_2 -Norm zugrundegelegt. Bei Verwendung anderer Normen ist eine nachträgliche Transformation mit einer geeigneten Diagonalmatrix erforderlich. Dabei bleibt die Überlaufstabilität erhalten. Das System ist allerdings nicht mehr rauschoptimal. Voraussetzung hierfür war nämlich, daß die Aussteuerungen Zustandsvariablen bzgl. der L_2 -Norm gleich sind. Das umskalierte Filter wird diese Bedingung i.a. höchstens näherungsweise erfüllen.

In Tabelle 3.1 sind noch einmal die $\underline{A}$ -Matrizen für die verschiedenen Fälle angegeben und denen der Normalform-Filter gegenübergestellt. Zur Vereinfachung der Schreibweise wird hier wieder die Bezeichnung $\underline{A}$ anstelle von $\underline{A}'$ verwendet, obwohl es sich um skalierte Systeme handelt.

Wie in Anhang A5 gezeigt wird, gehen die beiden Strukturen sogar ineinander über, d.h. $\underline{A}$ wird normal, wenn die Eigenwerte in die Nähe des Einheitskreises rücken. Dies gilt sowohl bei Skalierung nach der L_2 -Norm als auch bei Skalierung nach der L_∞ -Norm.

Außerdem entnimmt man aus Anhang A5 für den betrachteten Grenzfall, daß bei Gleichheit der L_∞ -Normen auch die L_2 -Normen gleich sind. Dieses Nebenergebnis läßt den Schluß zu, daß ein L_∞ -skaliertes Block zweiten Grades näherungsweise rauschoptimal ist, wenn die Polradien groß sind. Entsprechendes gilt für in Parallelform realisierte Filter höheren Grades. Die Aussage deckt sich im übrigen mit den Beobachtungen von Jackson, Lindgren und Kim /23/.

Ergänzend sei auf eine Arbeit von Mills, Mullis und Roberts /54/ verwiesen, worin die rauschoptimale Struktur der Normalform gegenübergestellt wird (Filtergrad 2). Das Schwergewicht liegt hier auf einem Vergleich der theoretischen Formeln für das Rundungsrauschen.

	Rauschoptimale Zustands- raum - Struktur	Normalform - Struktur
Polstellen konjugiert komplex $\lambda_{1,2} =$ $\sigma \pm j\omega$	$\underline{A} = \begin{bmatrix} \sigma & \sqrt{\frac{G_2}{G_1}} \omega \\ -\sqrt{\frac{G_1}{G_2}} \omega & \sigma \end{bmatrix}$ $= \underline{G}^{-1/2} \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} \underline{G}^{1/2}$ mit $\underline{G} = \text{Diag}\{G_1, G_2\}$ $ \underline{A} \underline{x} _{\underline{G}} = \sqrt{\sigma^2 + \omega^2} \underline{x}' _{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n$	$\underline{A} = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}$ $ \underline{A} \underline{x} _{\underline{E}} = \sqrt{\sigma^2 + \omega^2} \underline{x} _{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n$
Polstellen λ_1, λ_2 reell	<u>Fall 1:</u> $\underline{A} = \begin{bmatrix} \frac{\lambda_1 + \lambda_2}{2} & \sqrt{\frac{G_2}{G_1}} \frac{\lambda_1 - \lambda_2}{2} \\ \sqrt{\frac{G_1}{G_2}} \frac{\lambda_1 - \lambda_2}{2} & \frac{\lambda_1 + \lambda_2}{2} \end{bmatrix}$ $= \underline{G}^{-\frac{1}{2}} \begin{bmatrix} \frac{\lambda_1 + \lambda_2}{2} & \frac{\lambda_1 - \lambda_2}{2} \\ \frac{\lambda_1 - \lambda_2}{2} & \frac{\lambda_1 + \lambda_2}{2} \end{bmatrix} \underline{G}^{\frac{1}{2}}$ $ \underline{A} \underline{x} _{\underline{G}} \leq \max\{ \lambda_1 , \lambda_2 \} \cdot \underline{x} _{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n$ <u>Fall 2:</u> $\underline{A}$ symmetrisch $\rightarrow$ normal $ \underline{A} \underline{x} _{\underline{E}} \leq \max\{ \lambda_1 , \lambda_2 \} \cdot \underline{x} _{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n$	$\underline{A}$ symmetrisch $ \underline{A} \underline{x} _{\underline{E}} \leq \max\{ \lambda_1 , \lambda_2 \} \cdot \underline{x} _{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n$

Tabelle 3.1: Vergleich der $\underline{A}$ -Matrizen für Filtergrad 2

3.2.4 Zusammenfassung

Im vorliegenden Abschnitt wurden zwei überlaufstabile Zustandsraum-Filter, die Normalform-Struktur nach Barnes und die rauschoptimale Form nach Mullis und Roberts, beschrieben.

Im ersten Fall gewährleistet die Wahl einer normalen A-Matrix, daß die gemäß

$$||\underline{x}|| = \sqrt{\underline{x}^T \underline{G} \underline{x}} \quad (2-6)$$

definierte Norm des ZV-Vektors x bei verschwindender Erregung in jedem Filterschritt abnimmt. Aufgrund der Ergebnisse von Kapitel 2 ist daher das System inkremental pseudopassiv realisierbar, wenn als Überlaufcharakteristik eine Dreiecks- oder Sättigungskennlinie verwendet wird. Eine Besonderheit der Struktur besteht darin, daß bei der Norm-Definition des ZV-Vektors x die Diagonalmatrix G stets gleich der Einheitsmatrix gesetzt werden kann. Dies bedeutet aber gleichzeitig eine Einschränkung gegenüber dem allgemeinen Fall eines überlaufstabilen Filters. Verzichtet man auf die Forderung G = E, so ist z.B. eine nachträgliche Skalierung nach einer anderen als der zunächst verwendeten L_2 -Norm möglich. Der Spielraum kann aber auch zu einer weiteren Reduzierung des Rundungsrauschens genutzt werden. Dies führt zu der in Abschnitt 3.2.3 beschriebenen rauschoptimalen Form nach Mullis und Roberts. Allerdings ist die Überlaufstabilität hier nur für Filterblöcke zweiten Grades bzw. aus diesen aufgebaute Systeme höherer Ordnung gewährleistet. Es bestehen enge Verwandtschaften zu den Normalform-Strukturen zweiten Grades. So sind die A-Matrizen entweder gleich oder über eine einfache Ähnlichkeitstransformation mit einer reellen Diagonalmatrix miteinander verknüpft. Für Polstellen nahe dem Einheitskreis gehen die A-Matrizen beider Realisierungen wegen der Skalierungsbedingung sogar ineinander über.

Neben den erwähnten günstigen Quantisierungseigenschaften der angesprochenen Strukturen darf ein wesentlicher Nachteil nicht übersehen werden. Die Elemente der resultierenden System-Matrizen A, b, c^T, d sind nämlich i.a. ungleich Null oder Eins, so daß zur

Berechnung eines Filterausgangswertes $(n+1)^2$ Multiplikationen (n = Filtergrad) erforderlich sind. Das bedeutet für einen Block 2-ten Grades bereits 9 gegenüber 5 Multiplikationen bei der ersten kanonischen Form. Bei höhergradigen Normalform-Filtern sind die Verhältnisse noch wesentlich ungünstiger, da der Aufwand quadratisch mit der Ordnung n wächst. Es liegt daher nahe, auch bei dieser Struktur das System aus Blöcken zweiten Grades zu kombinieren. Eine andere mögliche Alternative besteht in einer segmentweisen Verarbeitung der Eingangsfolge nach Lingren und Zeman, wie an späterer Stelle (Kap. 6) erläutert wird.

3.3 Allgemeine Wellendigitalfilter

3.3.1 Vorbemerkungen

Hierbei handelt es sich um Strukturen, welche auf der Nachbildung zeitkontinuierlicher Filter beruhen. Die so gewonnenen Systeme werden als Wellendigitalfilter bezeichnet, weil als Signalgrößen die aus der Streuparametertheorie bekannten Wellen verwendet werden. Prinzipiell können beliebige zeitkontinuierliche Filter als Vorlage (Referenzfilter) herangezogen werden. Von besonderer Bedeutung sind aber die beidseitig resistiv abgeschlossenen verlustfreien Abzweigstrukturen, da sich durch einen geeigneten Entwurf deren geringe Toleranzempfindlichkeit auf die Koeffizienten des digitalen Systems übertragen läßt. Die Koeffizienten-Darstellung kann daher mit außerordentlich kurzer Wortbreite erfolgen, ein Vorteil, der allerdings nur dann zum Tragen kommt, wenn das Filter in Hardware realisiert wird.

Die Wellendigitalfilter sind darüber hinaus wegen ihrer garantierten Überlaufstabilität von Interesse.

In den vergangenen Jahren erschien eine Vielzahl von Publikationen zu diesen Strukturen, die im Einzelnen hier nicht aufgeführt werden sollen. Es sei lediglich auf die grundlegende Arbeit von Fett-

weis /2/, auf die Dissertationen von Sedlmeyer /27/ und Meerkötter /9/ sowie eine Übersichtsarbeit von Scheuermann /28/ verwiesen.

Da eine vollständige Diskussion der sehr umfangreichen Problematik im Rahmen der vorliegenden Arbeit nicht sinnvoll erscheint, sollen lediglich einige Grundgedanken erläutert werden. Die Darstellung orientiert sich im wesentlichen an /28/. Zusätzlich wird ein einfacher Nachweis für die Überlaufstabilität derartiger Strukturen erbracht, wobei die allgemeinen theoretischen Ergebnisse aus Abschnitt 2.3 zur Anwendung kommen. Schließlich sollen einige Eigenschaften der Systemmatrix A diskutiert werden.

3.3.2 Netzwerkbeschreibung durch Wellengrößen

Wir betrachten ein zeitkontinuierliches N-Tor mit normierten Torströmen i_v und Torspannungen u_v , $v = 1(1)N$, wie in Bild 3.1 dargestellt. Dem Tor v wird ferner ein Torleitwert G_v in Form einer geeignet zu wählenden positiven reellen Konstanten zugeordnet.

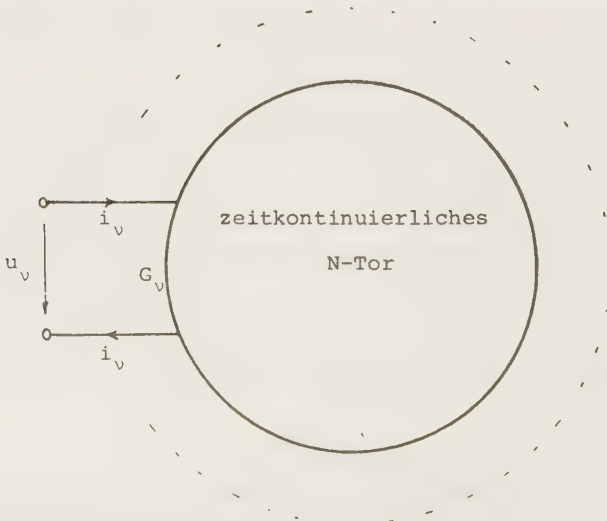


Bild 3.1: Zeitkontinuierliches N-Tor

Man definiert nun die Spannungswellen am Tor v zu

$$\begin{aligned} a_v &= u_v + i_v / G_v \\ b_v &= u_v - i_v / G_v, \end{aligned} \quad (3-33)$$

wobei hierunter die Augenblickswerte von einlaufender (a_v) und auslaufender Welle (b_v) zu verstehen sind (Bild 3.2).

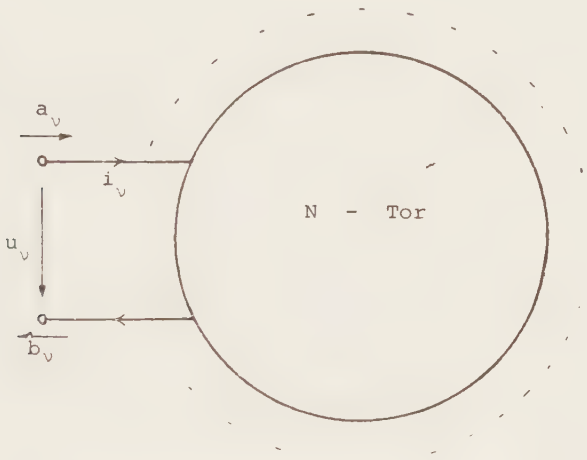


Bild 3.2: Zur Definition der Wellengrößen

Der Grund für die Verwendung derartiger Wellengrößen im Zusammenhang mit dem Entwurf digitaler Filter besteht darin, daß sich dann verzögerungsfreie Schleifen im Signalflußgraphen vermeiden lassen. Dies ist bekanntlich eine unabdingbare Voraussetzung für die Realisierung in digitaler Form.

3.3.3 Zusammenhang zwischen s- und z-Bereich

"s" sei die normierte komplexe Frequenzvariable der Laplace-Transformation, "z" die Variable der Z-Transformation zur Beschreibung zeitdiskreter Systeme und Signale. z und s werden nun mittels der bilinearen Transformation

$$s = \frac{z-1}{z+1} \quad (3-34)$$

zueinander in Beziehung gesetzt. Damit bildet man die imaginäre Achse der s-Ebene auf den Einheitskreis der z-Ebene eineindeutig ab. Mit $s = j\omega$ und $z = e^{j\Omega}$ erhält man den Zusammenhang

$$\omega = \tan \frac{\Omega}{2} \quad (3-35)$$

bzw. $\Omega = 2 \arctan \omega$

Diese Transformation wird u.a. beim Entwurf rekursiver digitaler Systeme mit vorgeschriebenem Verhalten im Frequenzbereich angewandt (z.B. /6/), wobei die Übertragungsfunktion $H(z)$ aus dem zuvor im s-Bereich bestimmten $\tilde{H}(s)$ gemäß

$$H(z) = \tilde{H} \left(\frac{z-1}{z+1} \right)$$

gewonnen wird. Demgegenüber wird bei der Herleitung eines Wellendigitalfilters die Transformation direkt auf die frequenzabhängigen Impedanzen des Referenzfilters angewandt, was aber bezüglich der Übertragungsfunktion zum gleichen Resultat führt. Über die spätere digitale Realisierung ist an dieser Stelle noch nichts ausgesagt.

3.3.4 Wellenbeziehungen elementarer Bausteine

Die Auswertung erfolgt aufgrund der Wellendefinitionen (3-33). Die Ergebnisse, welche in Tabelle 3.2 zusammengestellt sind, sollen an einigen grundsätzlichen Beispielen erläutert werden.

Ohmscher Widerstand

Wegen $i/G = u$ folgt

$$a = 2u, \quad b = 0,$$

wobei G gleich dem Widerstandsleitwert gesetzt wurde.

Widerstandsbehaftete Spannungsquelle

Da es sich um eine Quelle handelt, soll die Orientierung des Stromes sowie der Wellengrößen a und b umgekehrt werden. Dann ergibt sich $a = u_0$.

Offenbar ist a unabhängig von b durch die Quellspannung u_0 festgelegt. Somit stellt die Spannungsquelle auch eine Wellenquelle im Sinne obiger Definitionen dar.

Induktivität

Anstelle der Zeitsignale sollen hier die zugehörigen Laplace-Transformierten verwendet werden, was durch die Bezeichnung mit entsprechenden Großbuchstaben zum Ausdruck kommt. Mit $U = sLI$ und dem Torleitwert $G = \frac{1}{L}$ (normierte Größen) folgt

$$A = sLI + LI = I \cdot L [1+s]$$

$$B = sLI - LI = I \cdot L [s-1].$$

Durch Substitution von s gemäß der bilinearen Transformation (3-34) erhält man

$$A = I \cdot L \frac{2z}{z+1}, \quad B = -IL \frac{z}{z+1},$$

woraus sich der Zusammenhang zwischen A und B als

$$B = -\frac{1}{z} A$$

ergibt.

Die Induktivität läßt sich also durch eine im digitalen Fall besonders einfach zu realisierende Verzögerung um eine Abtastperiode nachbilden.

Element	Schaltbild	Strom-/ Spannungs- Beziehungen	Wellen- beziehungen	Wellen- fluß- diagramm
ohmscher Wider- stand		$u = i/G$	$b = 0$ Torleit- wert G	
Kurz- schluß		$u = 0$	$b = -a$ Torleit- wert G be- liebig	
Leer- lauf		$i = 0$	$b = a$ Torleit- wert G be- liebig	
Idealer Über- trager		$u_1 = \dot{u} \cdot u_2$ $i_2 = -\dot{u} \cdot i_1$	$b_1 = \dot{u} \cdot a_2$ $b_2 = \frac{1}{\dot{u}} \cdot a_1$ $\frac{G_2}{G_1} = \dot{u}^2$	
Gyrator		$u_1 = -i_2/G$ $u_2 = i_1/G$	$b_1 = -a_2$ $b_2 = a_1$ $G_1 = G_2 = G$	
3-Tor- zirkula- tor		$i_1/G = u_2 - u_3$ $i_2/G = u_3 - u_1$ $i_3/G = u_1 - u_2$	$b_1 = a_3$ $b_2 = a_1$ $b_3 = a_2$ $G_1 = G_2 = G_3 = G$	
Spannungs- quelle mit Wider- stand		$u = u_0 - i/G$	$a = u_0$ Torleit- wert G	

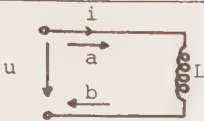
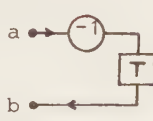
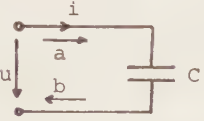
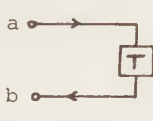
Element	Schaltbild	Strom-/ Spannungs- Beziehungen	Wellen- beziehungen	Wellen- fluß- diagramm
Induk- tivi- tät		$U = sLI$	$B = -\frac{1}{z} A$ Torleit- wert $G = 1/L$	
Kapazi- tät		$U = \frac{I}{sC}$	$B = \frac{1}{z} A$ Torleit- wert $G = C$	

Tabelle 3.2: Wellenbeziehungen elementarer Bausteine (L, C, G sind normierte pos. Konstanten). (Die Ergebnisse sind aus /28/ übernommen).

3.3.5 Verbindung der Tore zum Netzwerk

Eine direkte Verbindung zweier Tore im Wellenflußdiagramm ist i.a. nicht möglich, da aufgrund der unterschiedlichen Torleitwerte Reflexionen auftreten. Ihre Auswirkung auf die Wellenbeziehungen soll am Beispiel einer Parallelschaltung von Teilnetzwerken (Bild 3.3) erläutert werden. Gesucht ist hierbei der Zusammenhang zwischen einlaufenden Wellen a_μ und auslaufenden Wellen b_μ , $\mu = 1(1)N$. Mit Definition (3-33) gilt für die einzelnen Tore

$$a_\mu = u_\mu + i_\mu / G_\mu \quad (\text{I})$$

$$\mu = 1(1)N$$

$$b_\mu = u_\mu - i_\mu / G_\mu \quad (\text{II})$$

Ferner gilt gemäß den Kirchhoff'schen Gesetzen

$$u_1 = u_2 = \dots = u_N = u \quad (\text{III})$$

$$\sum_{v=1}^N i_v = 0 \quad (\text{IV})$$

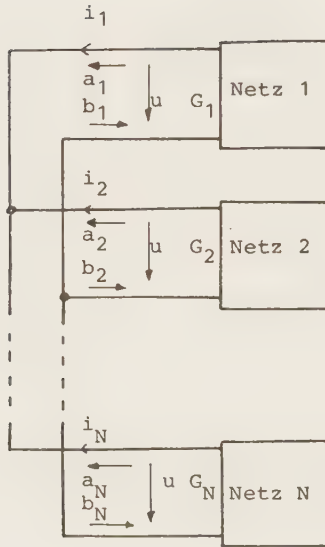


Bild 3.3: Parallelschaltung von Teilnetzwerken

Nach Zwischenrechnung erhält man aus (I) - (IV) das gewünschte Ergebnis.

$$b_{\mu} = -a_{\mu} + \frac{2}{\sum_{\kappa=1}^N G_{\kappa}} \sum_{v=1}^N G_v a_v$$

Zur Realisierung derartiger Beziehungen sind im Wellendigitalfilter zusätzliche Bausteine, die sogenannten "Adaptoren" erforderlich. In ihnen steckt der eigentliche arithmetische Aufwand des Systems.

3.3.6 Allgemeine Anmerkungen zur Realisierung

Ein Nachteil der Wellendigitalfilter besteht allgemein darin, daß kein modularer Aufbau möglich ist, wie das etwa bei üblichen Kaskaden- oder Parallelanordnungen von Blöcken zweiten Grades der Fall ist. Darüber hinaus ändert sich der digitale Aufbau grund-

sätzlich mit der Struktur des Referenzfilters. Schließlich ist keine gezielte Skalierung im Innern der Adaptoren möglich. Die Aussteuerung an den Adaptorausgängen läßt sich zwar beeinflussen, doch sind hierbei zusätzliche ideale Übertrager (d.h. deren digitale Realisierung) an den Ein- und Ausgängen der Adaptoren notwendig, deren Übersetzungsverhältnisse invers zueinander sind. (Am Analogrechner spricht man hier vom "Schnittprinzip"). Zur Vermeidung von Übersteuerungen im Innern schlägt z.B. Meerkötter /9/ eine lokale Wortlängenerhöhung vor, was aber bei Implementierung an einem Signalprozessor aufwendig ist.

3.3.7 Überlaufstabilität

Nach /9/ wird die notwendige Rundung bzw. Überlaufkorrektur an den Ausgängen der Adaptoren und idealen Übertrager ausgeführt. Für die theoretische Beschreibung wird angenommen, daß an der Verbindung zweier Tore ein 3-Tor-Zirkulator eingefügt wird, dessen freies Tor mit dem Quantisierer beschaltet wird (Bild 3.4).

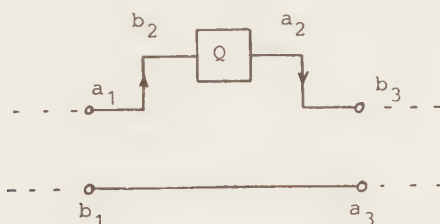


Bild 3.4: Quantisierung mit Hilfe eines 3-Tor-Zirkulators

Die Überlaufstabilität wurde bislang (z.B. /9/) mit der "Pseudopassivität" der einzelnen Bausteine eines Wellendigitalfilters begründet. Demgegenüber soll der Nachweis nun anhand des in Abschnitt 2.3 eingeführten allgemeinen Filtermodells erbracht werden.

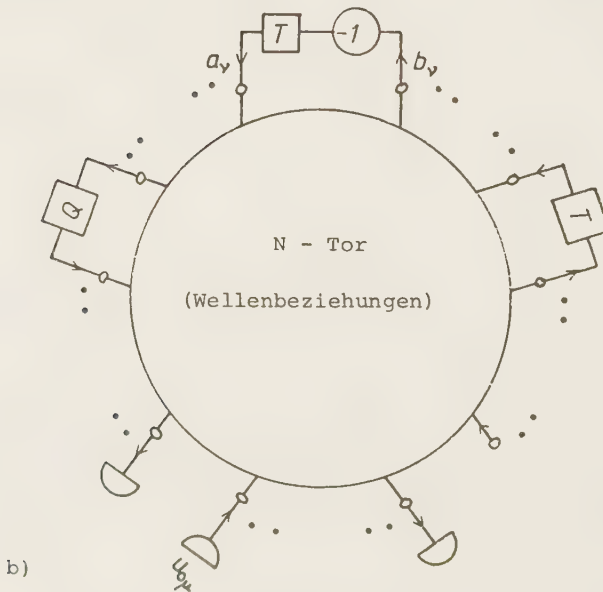
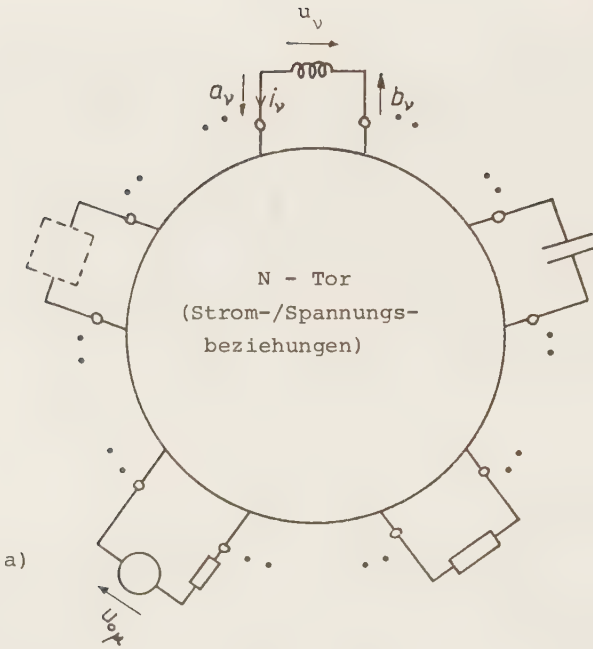


Bild 3.5: a) allgemeines Referenzfilter
b) abgeleitetes Wellendigitalfilter

Dazu betrachten wir das zeitkontinuierliche Filter als N-Tor gemäß Bild 3.5a. Die externen Tore seien mit allen vorkommenden reaktiven Elementen, Widerständen und widerstandsbehafteten Quellen beschaltet. Darüber hinaus treten nicht näher spezifizierte gestrichelt gezeichnete Elemente auf, die für die Quantisierung in der späteren digitalen Realisierung stehen. Die zugehörigen Anschlüsse können z.B. nach außen geführte Tore von 3-Tor-Zirkulatoren sein. Das Innere der Anordnung möge weder energiespeichernde noch energieverbrauchende Elemente enthalten. Zuge lassen sind dagegen ideale Verbindungen, ideale Übertrager, Gyratoren und Zirkulatoren. Daraus folgt offenbar, daß die Summe der an den externen Toren eingespeisten Augenblicksleistungen gleich Null ist. Faßt man die Torspannungen und Ströme in den Vektoren $\underline{u}$ und $\underline{i}$ zusammen, so gilt

$$\underline{u}^T \underline{i} = 0 \quad \forall t. \quad (3-36)$$

Mit den Vektoren $\underline{a}$ und $\underline{b}$ der zugehörigen Wellengrößen und der Diagonalmatrix $\underline{G}$ der Torleitwerte erhält man mit Definition (3-33)

$$\underline{u} = \frac{1}{2} (\underline{a} + \underline{b}), \quad \underline{i} = \frac{1}{2} \underline{G} (\underline{a} - \underline{b})$$

Damit geht (3-36) über in

$$\underline{a}^T \underline{G} \underline{a} - \underline{b}^T \underline{G} \underline{b} + \underline{b}^T \underline{G} \underline{a} - \underline{a}^T \underline{G} \underline{b} = 0. \quad (3-37)$$

Da die auftretenden Summenglieder skalare Größen sind, gilt

$$\underline{a}^T \underline{G} \underline{b} = (\underline{a}^T \underline{G} \underline{b})^T = \underline{b}^T \underline{G} \underline{a}, \text{ und mit (3-37) folgt}$$

$$\underline{b}^T \underline{G} \underline{b} = \underline{a}^T \underline{G} \underline{a} \quad \forall \underline{a} \in \mathbb{R}^N \quad (3-38)$$

Der Übergang vom Referenzfilter zum Wellendigitalfilter geschieht nun derart, daß die frequenzabhängigen Impedanzen der bilinearen Transformation unterworfen werden und man außerdem die Wellen als Signalgrößen verwendet (Bild 3.5b). Im vorliegenden Fall beeinflusst die bilineare Transformation nur die externe Beschaltung des N-Tores. Nachdem andererseits bei der Herleitung von (3-38)

keine Voraussetzungen bezüglich dieser externen Beschaltung getroffen wurden, gilt (3-38) auch im digitalen Fall.

Ein Vergleich von Bild 3.5b mit dem in Abschnitt 2.3.1 eingeführten allgemeinen Filtermodell zeigt, daß a und b offenbar bis auf das Vorzeichen einzelner Elemente von b mit den dort definierten Signalvektoren identisch sind. Somit besitzt aber wegen (3-38) das Wellendigitalfilter die Eigenschaft (2-24a). Nach Abschnitt 2.3 ist dies eine hinreichende Bedingung für die inkrementale Pseudopassivität, sofern als Überlaufcharakteristik eine Dreiecks- oder Sättigungskennlinie verwendet wird. Außerdem ist das System bei Betragsabschneiden frei von granularen Grenzyklen.

3.3.8 Eigenschaften der A-Matrix

Bezüglich der Überlaufstabilität ist die A-Matrix von untergeordneter Bedeutung, da die Quantisierer nicht nur an den Eingängen der Zustandsspeicher sondern an beliebigen Verbindungen zweier Tore auftreten können. Dennoch sind die Eigenschaften von A von Interesse, um die Wellendigitalfilter mit anderen überlaufstabilen Strukturen vergleichen zu können.

3.3.8.1 Allgemeine Wellendigitalfilter

Wir betrachten ein zeitkontinuierliches Referenzfilter als N-Tor gemäß Bild 3.6a, das im Innern ideale Verbindungen, ideale Übertrager, Gyratoren und Zirkulatoren enthalten darf. Die externen Tore seien mit den reaktiven Elementen und Widerständen beschaltet. Da letztlich nur die A-Matrix interessiert, wurden hierbei alle Quellen durch ihre Innenwiderstände ersetzt. Das zugehörige Wellendigitalfilter ist in Bild 3.6b dargestellt, wobei die Arithmetik als ideal angenommen wird. Die durch das N-Tor realisierten Wellenbeziehungen werden für das Referenzfilter und für das Wellendigitalfilter durch die gleiche reelle Streumatrix S gemäß

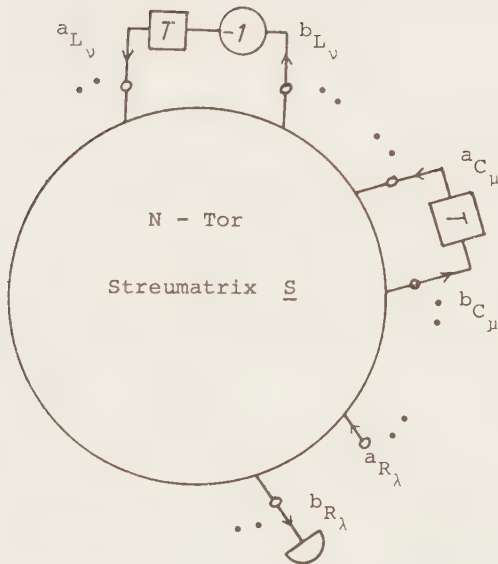
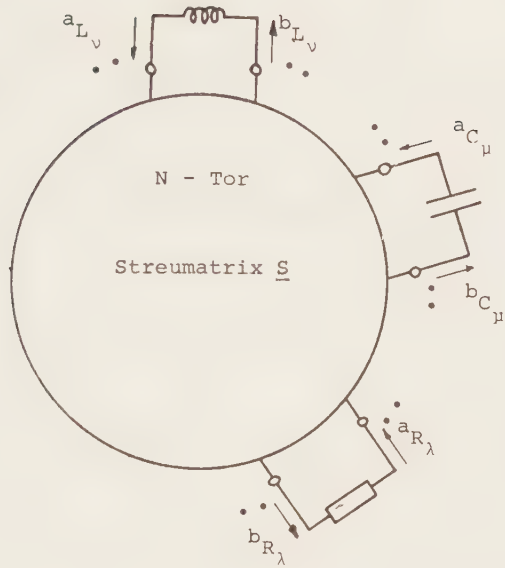


Bild 3.6: a) allgemeines Referenzfilter ohne Quellen
b) abgeleitetes Wellendigitalfilter mit idealer Arithmetik

$$\underline{b} = \underline{S} \underline{a} ; \quad \underline{S} : (N \times N)\text{-Matrix} \quad (3-39)$$

beschrieben /9/.

Wie bereits im vorangegangenen Abschnitt 3.3.7, Glg. (3-38), ausgeführt, gilt auch hier

$$\underline{b}^T \underline{G} \underline{b} = \underline{a}^T \underline{G} \underline{a} \quad \forall \underline{a} \in \mathbb{R}^N. \quad (3-40)$$

Mit (3-39) folgt

$$\begin{aligned} \underline{a}^T \underline{S}^T \underline{G} \underline{S} \underline{a} &= \underline{a}^T \underline{G} \underline{a} \quad \forall \underline{a} \in \mathbb{R}^N \\ \rightarrow \underline{a}^T \underline{G}^{\frac{1}{2}} \underline{G}^{-\frac{1}{2}} \underline{S}^T \underline{G}^{\frac{1}{2}} \underline{G}^{\frac{1}{2}} \underline{S} \underline{G}^{-\frac{1}{2}} \underline{G}^{\frac{1}{2}} \underline{a} &= \underline{a}^T \underline{G}^{\frac{1}{2}} \underline{G}^{\frac{1}{2}} \underline{a} \\ \rightarrow \left[\left(\underline{G}^{\frac{1}{2}} \underline{S} \underline{G}^{-\frac{1}{2}} \right) \left(\underline{G}^{\frac{1}{2}} \underline{a} \right) \right]^T \left[\left(\underline{G}^{\frac{1}{2}} \underline{S} \underline{G}^{-\frac{1}{2}} \right) \left(\underline{G}^{\frac{1}{2}} \underline{a} \right) \right] &= \left[\underline{G}^{\frac{1}{2}} \underline{a} \right]^T \left[\underline{G}^{\frac{1}{2}} \underline{a} \right] \\ \forall \left[\underline{G}^{\frac{1}{2}} \underline{a} \right] &\in \mathbb{R}^N \quad (3-41) \end{aligned}$$

Demnach muß die als

$$\tilde{\underline{S}} := \underline{G}^{1/2} \underline{S} \underline{G}^{-1/2} \quad (3-42)$$

definierte Matrix orthogonal sein.

Wir unterteilen nun die Vektoren $\underline{a}$ und $\underline{b}$ entsprechend der Zuordnung zu den Induktivitäten, Kapazitäten und Widerständen des Referenzfilters:

$$\underline{a} = \begin{bmatrix} \underline{a}_L \\ \underline{a}_C \\ \underline{a}_R \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} \underline{b}_L \\ \underline{b}_C \\ \underline{b}_R \end{bmatrix} \quad (3-43)$$

$\underline{a}_L, \underline{b}_L$: $(n_1 \times 1)$ Spaltenvektoren

$$n_1 + n_2 = n$$

$\underline{a}_C, \underline{b}_C$: $(n_2 \times 1)$ Spaltenvektoren

$\underline{a}_R, \underline{b}_R$: $(n_3 \times 1)$ Spaltenvektoren;

$$n_1 + n_2 + n_3 = N$$

Aufgrund der Wellenbeziehung für ohmsche Widerstände gilt

$$\underline{a}_R = \underline{0} \quad (3-44)$$

Somit ergibt sich der Zustandsvektor der digitalen Realisierung gemäß Bild 3.6 zu

$$\underline{x}(k) = \begin{bmatrix} \underline{a}_L(k) \\ \underline{a}_C(k) \end{bmatrix}, \quad \underline{x}(k+1) = \begin{bmatrix} -\underline{b}_L(k) \\ \underline{b}_C(k) \end{bmatrix} \quad (3-45)$$

Zusammen mit der Streumatrix-Darstellung (3-39) folgt

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{b}_R(k) \end{bmatrix} = \begin{bmatrix} -\underline{E}_{n_1} & & \\ & \underline{E}_{n_2} & \\ & & \underline{E}_{n_3} \end{bmatrix} \cdot \underline{S} \cdot \begin{bmatrix} \underline{x}(k) \\ \underline{0}_{n_3} \end{bmatrix} \quad (3-46)$$

$\underline{E}_{n_v} = (n_v \times n_v)$ -Einheitsmatrix

$\underline{0}_{n_3} = (n_3 \times 1)$ -Nullvektor.

Durch linksseitige Multiplikation mit der Diagonalmatrix $\underline{G}^{1/2}$ und einfache Umformung gewinnt man hieraus

$$\underline{G}^{1/2} \begin{bmatrix} \underline{x}_1(k+1) \\ \underline{b}_R(k) \end{bmatrix} = \begin{bmatrix} -\underline{E}_{n_1} & & \\ & \underline{E}_{n_2} & \\ & & \underline{E}_{n_3} \end{bmatrix} \underbrace{\underline{G}^{1/2} \underline{S} \underline{G}^{-1/2}}_{= \underline{\hat{S}}} \cdot \underline{G}^{1/2} \begin{bmatrix} \underline{x}(k) \\ \underline{0}_{n_3} \end{bmatrix} \quad (3-47)$$

$:= \underline{\hat{S}}$

Da die bereits in (3-42) angegebene Matrix $\underline{\hat{S}}$ orthogonal ist, muß auch $\underline{\hat{S}}$ orthogonal sein. Somit besitzen die Vektoren

$$\underline{G}^{1/2} \begin{bmatrix} \underline{x}(k) \\ \underline{0}_{n_3} \end{bmatrix}$$

und

$$\underline{G}^{1/2} \begin{bmatrix} \underline{x}(k+1) \\ \underline{b}_R(k) \end{bmatrix}$$

die gleiche euklidische Norm. Bezeichnen wir nun mit $\underline{G}_1$ die Diagonalmatrix der ersten n Elemente von $\underline{G}$, ($n=n_1+n_2$), dann gilt offenbar

$$||\underline{G}_1^{1/2} \underline{x}(k+1)||_{\underline{E}} \leq ||\underline{G}_1^{1/2} \underline{x}(k)||_{\underline{E}} \quad \forall \underline{x}(k) \in \mathbb{R}^n \quad (3-48)$$

$$||\dots||_{\underline{E}} = \text{euklidische Norm}.$$

Das Gleichheitszeichen ist denkbar, wenn die Operation (3-47) ein $\underline{b}_R(k) = \underline{0}$ liefert.

(3-48) ist nun gleichbedeutend mit

$$\underline{x}^T(k+1) \underline{G}_1 \underline{x}(k+1) \leq \underline{x}^T(k) \underline{G}_1 \underline{x}(k) \quad \forall \underline{x}(k) \in \mathbb{R}^n \quad (3-49)$$

$$\text{bzw.} \quad ||\underline{x}(k+1)||_{\underline{G}_1} \leq ||\underline{x}(k)||_{\underline{G}_1}$$

Da andererseits $\underline{x}(k+1) = \underline{A} \underline{x}(k)$ gilt, folgt

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n. \quad (3-50)$$

Dieses Ergebnis ist nicht überraschend. Es bestätigt nur die bereits in Abschnitt 2.3.4 gefundene Aussage, daß nämlich aus der Stabilitätsbedingung (2-24), die ja von den Wellendigitalfiltern erfüllt wird, die Eigenschaft $||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1}$ folgt.

Die vergleichsweise aufwendige Herleitung im vorliegenden Abschnitt gestattet allerdings noch weitergehende Aussagen.

Im Fall der in Abschnitt 3.2 beschriebenen Zustandsraum-Strukturen galt bei geeigneter Wahl einer positiven Diagonalmatrix $\underline{G}_1$

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq \max_{\forall} \{|\lambda_v|\} ||\underline{x}||_{\underline{G}_1} < ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n, \quad \underline{x} \neq \underline{0}$$

Unter der Voraussetzung linearer Stabilität mit $|\lambda_v| < 1$, war also stets $\|\underline{A} \underline{x}\|_{\underline{G}_1}$ echt kleiner $\|\underline{x}\|_{\underline{G}_1}$. Demgegenüber wird sich bei den Wellendigitalfiltern zeigen, daß in (3-49), (3-50) unter gewissen Bedingungen auch das Gleichheitszeichen für $\underline{x} \neq \underline{0}$ auftreten kann. Dies ist wegen (3-47) und der Orthogonalität von $\hat{\underline{S}}$ genau dann möglich, wenn die Gleichung

$$\begin{bmatrix} \hat{\underline{x}}(k+1) \\ \underline{0}_{n_3} \end{bmatrix} = \hat{\underline{S}} \begin{bmatrix} \hat{\underline{x}}(k) \\ \underline{0}_{n_3} \end{bmatrix} \quad (3-52)$$

mit $\hat{\underline{x}}(k) = \underline{G}_1^{1/2} \underline{x}(k)$

für spezielle $\hat{\underline{x}}(k) \in \mathbb{R}^n$, $\hat{\underline{x}}(k) \neq \underline{0}$, erfüllt werden kann. Die rechte Seite von (3-52) stellt nun eine Linearkombination der ersten n orthonormalen Spaltenvektoren $\hat{\underline{S}}_v$, $v = 1(1)n$, von $\hat{\underline{S}}$ dar, wobei die Gewichtung mit den Komponenten von $\hat{\underline{x}}(k)$ erfolgt. Alle $\hat{\underline{S}}_v$, $v = 1(1)N$, bilden eine Orthonormalbasis im $\mathbb{R}^N$. Die ersten $\hat{\underline{S}}_v$, $v = 1(1)n$, sind folglich orthonormale Basisvektoren eines n -dimensionalen Unterraumes UR_1 . Die rechte Seite von (3-52) liefert die Elemente dieses Unterraumes. Andererseits sind die Vektoren

$$\begin{bmatrix} \hat{\underline{x}}(k+1) \\ \underline{0}_{n_3} \end{bmatrix} \quad \text{Elemente eines ebenfalls } n\text{-dimensionalen Unterraumes } UR_2.$$

Gleichung (3-52) ist demnach für spezielle $\hat{\underline{x}}(k)$ erfüllbar, wenn UR_1 und UR_2 eine von Null verschiedene Schnittmenge besitzen, d.h. wenn $\dim[UR_1 \cap UR_2] > 0$ ist. Der Fall $\dim[UR_1 \cap UR_2] = 0$ liefert als Schnittmenge den Punkt $\begin{bmatrix} \underline{0}_n \\ \underline{0}_{n_3} \end{bmatrix}$, wofür (3-52) natürlich stets erfüllt ist. Nach /38/ gilt nun

$$\dim[UR_1 \cap UR_2] = \dim[UR_1] + \dim[UR_2] - \dim[UR_1 + UR_2]. \quad (3-53)$$

Wegen $\dim|UR_1| = \dim|UR_2| = n$ und $\dim|UR_1 + UR_2| \leq N$ folgt

$$\dim[UR_1 \cap UR_2] \geq 2n - N, \quad (3-54)$$

und wir erhalten als Ergebnis die Aussage:

Für $n > \frac{N}{2}$ kann die Gleichung (3-52) für spezielle $\hat{x}(k) \in \mathbb{R}^n$, $\hat{x}(k) \neq \underline{0}$ erfüllt werden und es gilt $||\underline{A} \underline{x}||_{\underline{G}_1} = ||\underline{x}||_{\underline{G}_1}$ mit $\hat{x} = \underline{G}_1^{-1/2} \underline{x}$. Dabei ist n gleich der Anzahl der reaktiven Elemente (L, C) und N gleich der Torzahl des Referenzfilters nach Bild 3.6a.

Es sei vermerkt, daß die Bedingung $n > \frac{N}{2}$ hinreichend, aber i.a. nicht notwendig ist. Eine genauere Aussage ist mit (3-53) möglich, wenn im Einzelfall die tatsächliche Dimension der Vereinigungsmenge $[UR_1 + UR_2]$ bekannt ist.

Aufgrund obiger Überlegungen ist allerdings auch für $n > \frac{N}{2}$ nicht auszuschließen, daß eine andere Diagonalmatrix $\underline{G}_2$ existiert, für die $||\underline{A} \underline{x}||_{\underline{G}_2} < ||\underline{x}||_{\underline{G}_2} \quad \forall \underline{x} \neq \underline{0}$ gilt, das Gleichheitszeichen also nicht auftreten kann. Im Fall eines RLCÜ-Referenzfilters wird sich jedoch zeigen, daß diese Möglichkeit nicht besteht.

Wir wollen schließlich noch die $\underline{A}$ -Matrix explizit in Abhängigkeit von der Streumatrix $\underline{S}$ angeben. Wegen (3-46) gilt offenbar

$$\begin{aligned} \begin{bmatrix} \underline{x}(k+1) \\ \underline{0}_{-n_3} \end{bmatrix} &= \underbrace{\begin{bmatrix} -\underline{E}_{n_1} & \underline{E}_{n_2} & \\ & & \underline{0}_{-n_3, n_3} \end{bmatrix} \underline{S} \begin{bmatrix} \underline{E}_{n_1} & \underline{E}_{n_2} & \\ & & \underline{0}_{-n_3, n_3} \end{bmatrix}} \\ &= \begin{bmatrix} \underline{A} & \underline{0}_{-n, n_3} \\ \underline{0}_{-n_3, n} & \underline{0}_{-n_3, n_3} \end{bmatrix} \end{aligned} \quad (3-55)$$

$\underline{0}_{-n_\mu, n_\nu} = (n_\mu \times n_\nu)$ -Nullmatrix

$$n = n_1 + n_2, \quad N = n + n_3 = n_1 + n_2 + n_3$$

und nach linksseitiger Multiplikation mit $\underline{G}^{1/2}$

$$\begin{aligned} \underline{G}^{\frac{1}{2}} \begin{bmatrix} \underline{x}(k+1) \\ \underline{0}_{n_3} \end{bmatrix} &= \begin{bmatrix} \underline{E}_{n_1} & & \\ & \underline{E}_{n_2} & \\ & & \underline{0}_{n_3, n_3} \end{bmatrix} \underbrace{\begin{bmatrix} -\underline{E}_{n_1} & & \\ & \underline{E}_{n_2} & \\ & & \underline{E}_{n_3} \end{bmatrix} \begin{bmatrix} \underline{G}^{\frac{1}{2}} \underline{S} & \underline{G} \\ \underline{G}^{\frac{1}{2}} \underline{S} & \underline{G} \\ \underline{G}^{\frac{1}{2}} \underline{S} & \underline{G} \end{bmatrix} \begin{bmatrix} \underline{E}_{n_1} & & \\ & \underline{E}_{n_2} & \\ & & \underline{0}_{n_3, n_3} \end{bmatrix}}_{= \hat{\underline{S}}} \underline{G}^{\frac{1}{2}} \begin{bmatrix} \underline{x}(k) \\ \underline{0}_{n_3} \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \underline{G}^{\frac{1}{2}} \underline{A} & \underline{0}_{n, n_3} \\ \underline{0}_{n_3, n} & \underline{0}_{n_3, n_3} \end{bmatrix} \underline{G}^{-\frac{1}{2}}}_{= \hat{\underline{A}}} \begin{bmatrix} \underline{G}^{\frac{1}{2}} \underline{A} & \underline{0}_{n, n_3} \\ \underline{0}_{n_3, n} & \underline{0}_{n_3, n_3} \end{bmatrix} \underline{G}^{-\frac{1}{2}} = \begin{bmatrix} \underline{G}^{\frac{1}{2}} \underline{A} & \underline{0}_{n, n_3} \\ \underline{0}_{n_3, n} & \underline{0}_{n_3, n_3} \end{bmatrix} \underline{G}^{-\frac{1}{2}} \end{aligned}$$

(3-56)

Die Matrix

$$\hat{\underline{A}} := \underline{G}_1^{1/2} \underline{A} \underline{G}_1^{-1/2} \quad (3-57)$$

ist also stets Untermatrix einer Orthogonalmatrix $\hat{\underline{S}}$. Es wird sich an späterer Stelle zeigen, daß diese Eigenschaft sehr viel allgemeiner für alle betrachteten Überlaufstabilen Strukturen gilt und letztlich äquivalent zu der in Abschnitt 2.2.4 bzgl. der $\underline{A}$ -Matrix formulierten Stabilitätsbedingung ist.

3.3.8.2 Aus einem RLCU-Referenzfilter abgeleitetes System

Wir legen wieder ein Referenzfilter nach Bild 3.6a zugrunde, in dessen Innerem jedoch nur ideale Verbindungen und ideale Übertrager zugelassen sind. Nach /9/ ist die Streumatrix $\underline{S}$ in diesem Fall selbstinvers (involutorisch) mit

$$\underline{S} = \underline{S}^{-1} \quad (3-58)$$

Daraus folgt für $\tilde{\underline{S}} = \underline{G}^{1/2} \underline{S} \underline{G}^{-1/2}$:

$$\tilde{\underline{S}} \tilde{\underline{S}} = \underline{G}^{1/2} \underline{S} \underline{G}^{-1/2} \underline{G}^{1/2} \underline{S} \underline{G}^{-1/2} = \underline{E} \rightarrow \tilde{\underline{S}} = \tilde{\underline{S}}^{-1}. \quad (3-59)$$

Da nach den Überlegungen des vorangegangenen Abschnittes $\tilde{\underline{S}}$ orthogonal ist, folgt ferner:

$$\tilde{\underline{S}} = \tilde{\underline{S}}^{-1} = \tilde{\underline{S}}^T \quad (3-60)$$

Wir definieren nun $\tilde{\underline{A}}$ als nordwestliche ($n \times n$)-Untermatrix von $\tilde{\underline{S}}$. $\tilde{\underline{A}}$ ist somit ebenfalls symmetrisch. Bei Beachtung von (3-56) erhält man den Zusammenhang mit der Systemmatrix $\underline{A}$ als

$$\tilde{\underline{A}} = \begin{bmatrix} -\underline{E}_{n_1} & \\ & \underline{E}_{n_2} \end{bmatrix} \underline{G}_1^{1/2} \underline{A} \underline{G}_1^{-1/2}. \quad (3-61)$$

Damit ergibt sich der im folgenden benötigte Zusammenhang

$$\begin{aligned} ||\tilde{\underline{A}}(\underline{G}_1^{1/2} \underline{x})||_{\underline{E}}^2 &= \underline{x}^T \underline{G}_1^{1/2} \tilde{\underline{A}}^T \tilde{\underline{A}} \underline{G}_1^{1/2} \underline{x} \\ &= \underline{x}^T \underline{G}_1^{1/2} \underline{G}_1^{-1/2} \underline{A}^T \underline{G}_1^{1/2} \underline{G}_1^{1/2} \underline{A} \underline{G}_1^{-1/2} \underline{G}_1^{1/2} \underline{x} \\ &= \underline{x}^T \underline{A}^T \underline{G}_1 \underline{A} \underline{x} = ||\underline{A} \underline{x}||_{\underline{G}_1}^2. \end{aligned} \quad (3-62)$$

Die Beziehung

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n \quad (3-63a)$$

ist somit äquivalent zu

$$||\tilde{\underline{A}}(\underline{G}_1^{1/2} \underline{x})||_{\underline{E}} \leq ||\underline{G}_1^{1/2} \underline{x}||_{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n. \quad (3-63b)$$

Nach Abschnitt 3.3.8.1 ist das Gleichheitszeichen für spezielle $\underline{x} \neq \underline{0}$ möglich, wenn $n > \frac{N}{2}$ gilt. Diese Bedingung ist erfüllt, wenn im Referenzfilter die Zahl n der Reaktanzen die Anzahl der Widerstände übersteigt. Ein wichtiges Beispiel sind die beidseitig

ohm'sch abgeschlossenen LC-Abszweigstrukturen.

Es soll nun gezeigt werden, daß dann auch keine andere positive Diagonalmatrix $\underline{G}_2$ existiert, derart daß in (3-63a,b) das Gleichheitszeichen für $\underline{x} \neq \underline{0}$ ausgeschlossen ist. Um das zu beweisen, gehen wir von einem $\underline{x}_0 \neq \underline{0}$ aus, für das in (3-63a,b) das Gleichheitszeichen gilt, und zeigen, daß bei Verwendung einer beliebigen anderen Diagonalmatrix zur Normierung das Gleichheitszeichen erhalten bleibt.

Wir nehmen also an, daß für $\underline{x} = \underline{x}_0 \neq \underline{0}$

$$||\tilde{\underline{A}}(\underline{G}_1)^{1/2} \underline{x}_0||_{\underline{E}} = ||\underline{G}_1^{1/2} \underline{x}_0||_{\underline{E}} \quad (3-64)$$

gilt. Da $\tilde{\underline{A}}$ symmetrisch und somit normal ist, gilt die Abschätzung

$$||\tilde{\underline{A}}(\underline{G}_1)^{1/2} \underline{x}_0||_{\underline{E}} \leq \max_v \{|\tilde{\lambda}_v|\} \cdot ||\underline{G}_1^{1/2} \underline{x}_0||_{\underline{E}} \quad (3-65)$$

wobei mit $\tilde{\lambda}_v$ die Eigenwerte von $\tilde{\underline{A}}$ bezeichnet werden. Die $\tilde{\lambda}_v$ sind i.a. nicht mit den Eigenwerten von $\underline{A}$ identisch.

Damit nun (3-64) erfüllt werden kann, muß

$$|\tilde{\lambda}_{\max}| := \max_v \{|\tilde{\lambda}_v|\} = 1 \quad (3-66)$$

gelten. Wegen der Symmetrie von $\tilde{\underline{A}}$ sind die $\tilde{\lambda}_v$, $v = 1(1)n$, reell und es folgt

$$\tilde{\lambda}_{\max} = \pm 1 \quad (3-67)$$

Wählen wir $\underline{G}_1^{1/2} \underline{x}_0$ als reellen Eigenvektor zum Eigenwert $\tilde{\lambda}_{\max}$, dann gilt

$$\tilde{\underline{A}} [\underline{G}_1^{1/2} \underline{x}_0] = \pm \underline{G}_1^{1/2} \underline{x}_0 \quad (3-68)$$

Sei nun $\underline{G}_2$ eine beliebige positive $(n \times n)$ Diagonalmatrix, so folgt

$$\tilde{\underline{A}} \underline{G}_1^{1/2} \underline{G}_2^{-1/2} \underline{G}_2^{1/2} \underline{x}_0 = \pm \underline{G}_1^{1/2} \underline{x}_0 \quad (3-69)$$

Nach linksseitiger Multiplikation mit $\underline{G}_2^{1/2} \underline{G}_1^{-1/2}$ erhält man hieraus

$$\underbrace{\underline{G}_2^{1/2} \underline{G}_1^{-1/2} \tilde{\underline{A}} \underline{G}_1^{1/2} \underline{G}_2^{-1/2} \underline{G}_2^{1/2}}_{:= \tilde{\underline{A}}} \underline{x}_0 = \pm \underline{G}_2^{1/2} \underline{x}_0, \quad (3-69)$$

und es folgt

$$|| \tilde{\underline{A}} \underline{G}_2^{1/2} \underline{x}_0 ||_{\underline{E}} = || \underline{G}_2^{1/2} \underline{x}_0 ||_{\underline{E}} \quad (3-70a)$$

bzw.

$$\underline{x}_0^T \underline{G}_2^{1/2} \tilde{\underline{A}}^T \tilde{\underline{A}} \underline{G}_2^{1/2} \underline{x}_0 = \underline{x}_0^T \underline{G}_2^{1/2} \underline{G}_2^{1/2} \underline{x}_0. \quad (3-70b)$$

$\tilde{\underline{A}}$ läßt sich unter Beachtung von (3-61) als

$$\tilde{\underline{A}} = \begin{bmatrix} -\underline{E}_{n_1} & \\ & \underline{E}_{n_2} \end{bmatrix} \underline{G}_2^{1/2} \underline{A} \underline{G}_2^{-1/2}$$

bestimmen. Nach Einsetzen in (3-70b) ergibt sich

$$\underline{x}_0^T \underline{A}^T \underline{G}_2 \underline{A} \underline{x}_0 = \underline{x}_0^T \underline{G}_2 \underline{x}_0 \quad (3-71a)$$

bzw.

$$|| \underline{A} \underline{x}_0 ||_{\underline{G}_2} = || \underline{x}_0 ||_{\underline{G}_2} \quad (3-71b)$$

Damit ist der Beweis für obige Behauptung erbracht.

Die Ergebnisse des vorliegenden Unterabschnittes sollen noch einmal kurz zusammengefaßt werden:

1. Die Matrix

$$\tilde{\underline{A}} = \begin{bmatrix} -\underline{E}_{n_1} & \\ & \underline{E}_{n_2} \end{bmatrix} \underline{G}_1^{1/2} \underline{A} \underline{G}_1^{-1/2} \quad (3-72)$$

ist symmetrisch. $\underline{A}$ ist hierbei die Systemmatrix des Wellendigitalfilters. $\underline{G}_1$ ist eine n-dim. positive Diagonalmatrix. Ihre Hauptdiagonalelemente sind die zu den Reaktanzen des Re-

ferenzfilters gehörenden Torleitwerte.

Bei Verwendung von $\underline{G}_1$ zur Definition der Vektornorm

$$||\underline{x}||_{\underline{G}_1} = \sqrt{\underline{x}^T \underline{G}_1 \underline{x}} \text{ gilt stets}$$

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n. \quad (3-73)$$

Unter der Voraussetzung, daß im Referenzfilter die Zahl der reaktiven Elemente größer ist als die der Widerstände, existieren immer spezielle Vektoren $\underline{x} = \underline{x}_0 \in \mathbb{R}^n$, $\underline{x}_0 \neq \underline{0}$, für die das Gleichheitszeichen mit

$$||\underline{A} \underline{x}_0||_{\underline{G}_1} = ||\underline{x}_0||_{\underline{G}_1} \quad (3-74)$$

erfüllt ist.

Ferner existiert auch keine andere positive Diagonalmatrix $\underline{G}_2$, für die

$$||\underline{A} \underline{x}||_{\underline{G}_2} < ||\underline{x}||_{\underline{G}_2} \quad \forall ||\underline{x}|| \in \mathbb{R}^n, \underline{x} \neq \underline{0}$$

gilt, das Gleichheitszeichen also ausgeschlossen ist.

Hierin unterscheidet sich diese Art der Wellendigitalfilter von den in Abschnitt 3.2 beschriebenen Zustandsraum-Strukturen, für die bei geeigneter Wahl einer Diagonalmatrix $\underline{G}_1$ stets

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq \max_v \{|\lambda_v|\} \cdot ||\underline{x}||_{\underline{G}_1} < ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \neq \underline{0}$$

gilt. λ_v bezeichnet die Eigenwerte von $\underline{A}$.

Insbesondere kann deshalb die $\underline{A}$ -Matrix dieser Wellendigitalfilter weder normal sein noch durch Ähnlichkeitstransformation mit einer Diagonalmatrix normal gemacht werden.

3.3.8.3 Aus einem RLÜ- oder RCÜ-Referenzfilter abgeleitetes System

Wie bereits im vorhergehenden Abschnitt erläutert, ist die Matrix

$$\tilde{\underline{A}} = \begin{bmatrix} -\underline{E}_{n_1} & \\ & \underline{E}_{n_2} \end{bmatrix} \underbrace{\underline{G}_1^{1/2} \underline{A} \underline{G}_1^{-1/2}}_{= \hat{\underline{A}}} \quad (3-61)$$

symmetrisch und somit normal. n_1 und n_2 bezeichnen die Zahl der Induktivitäten bzw. Kapazitäten des Referenzfilters. Besitzt das System nur kapazitive oder nur induktive Reaktanzen, so folgt aus (3-61) offenbar

$$\tilde{\underline{A}} = \hat{\underline{A}} \quad \text{oder} \quad \tilde{\underline{A}} = -\hat{\underline{A}}. \quad (3-76)$$

In beiden Fällen ist mit $\tilde{\underline{A}}$ auch die Matrix $\hat{\underline{A}}$ normal. Es folgt

$$||\hat{\underline{A}} \underline{G}_1^{1/2} \underline{x}||_{\underline{E}} \leq \max_v \{ |\lambda_v| \} \cdot ||\underline{G}_1^{1/2} \underline{x}||_{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n.$$

bzw.

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq \max_v \{ |\lambda_v| \} \cdot ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n \quad (3-77)$$

λ_v , $v = 1(1)n$, sind die Eigenwerte von $\hat{\underline{A}}$ bzw. $\underline{A}$.

Die $\underline{A}$ -Matrix der hier behandelten Filter läßt sich also durch Transformation mit einer geeigneten Diagonalmatrix normal machen. Somit ergibt sich zumindest für diesen Spezialfall eines Wellendigitalfilters eine gewisse Verwandtschaft zu den antwortstabilen Zustandsraum-Strukturen nach Mullis/Roberts und Barnes.

3.4 Leiterstrukturen nach Markel und Gray

3.4.1 Beschreibung der Strukturen

Diese aus der Literatur als "Lattice"-, "Ladder"- oder "Orthogonal Polynomial"-Filter bekannten Strukturen wurden von Markel und Gray in /31-35/ vorgeschlagen. Ihr grundsätzlicher Aufbau ist in Bild 3.7 dargestellt, wobei die Teil-Zweitore in ihrer allgemeinsten Form gemäß Bild 3.8 beschreibbar sind. Der Filterentwurf, welcher in /31/ ausführlich erläutert wird, liefert u. a. einen Satz Koeffizienten r_μ und führt auf eine Basisstruktur mit

$$\alpha_\mu = 1, \beta_\mu = -r_\mu, \gamma_\mu = 1 - r_\mu^2, \delta_\mu = r_\mu,$$

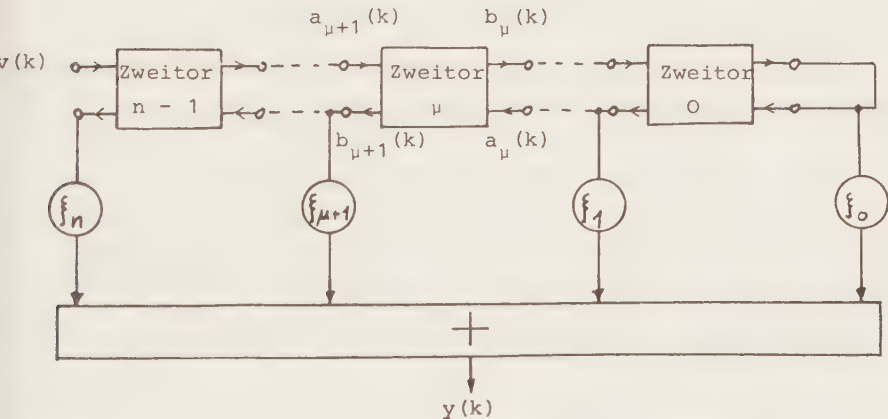


Bild 3.7: Grundsätzlicher Aufbau der Leiterstruktur nach Markel und Gray

wobei für im linearen Fall stabile Systeme $|r_\mu| < 1$, $\mu = 0(1)n-1$, gilt. Damit vereinfacht sich das o.a. allgemeine Zweitor zu der in Bild 3.9 gegebenen Anordnung, welche als "Zwei-Multiplizierer-Form" bekannt ist. Zu ihrer Kennzeichnung werden gestrichelte Signalgrößen verwendet. Hieraus lassen sich modifizierte Strukturen ableiten, wenn man alle Eingänge mit einem

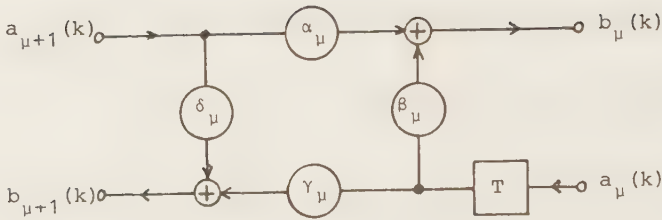


Bild 3.8: Allgemeine Form der Teil-Zweitore aus Bild 3.4

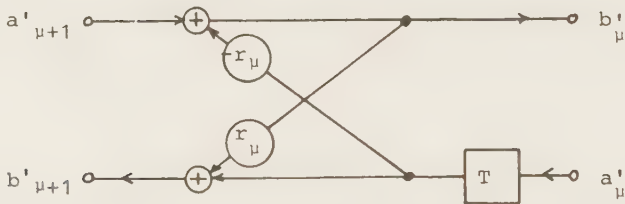


Bild 3.9: Teil-Zweitore in der Zwei-Multiplizierer-Form

Faktor π_μ multipliziert und dies durch Multiplikation mit $\frac{1}{\pi_\mu}$ an den Ausgängen kompensiert (Bild 3.10). Offenbar ändert sich das Verhalten des Zweitores nach außen hin nicht. Die Zusammenfassung der in der Kaskade jeweils aufeinanderfolgenden Multiplikationen liefert zusammen mit der Basis-Struktur nach Bild 3.9 die modifizierte Form gemäß Bild 3.11. Bezüglich der allgemeinen

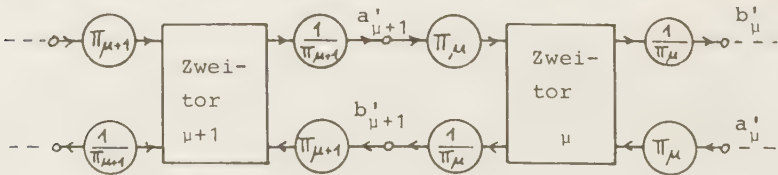


Bild 3.10: Zur Herleitung modifizierter Teil-Zweitore

Beschreibung (Bild 3.8) gilt nun

$$\alpha_\mu = \frac{\pi_\mu}{\pi_{\mu+1}}, \quad \beta_\mu = -r_\mu, \quad \gamma_\mu = [1 - r_\mu^2] \frac{\pi_{\mu+1}}{\pi_\mu}, \quad \delta_\mu = r_\mu \quad (3-78)$$

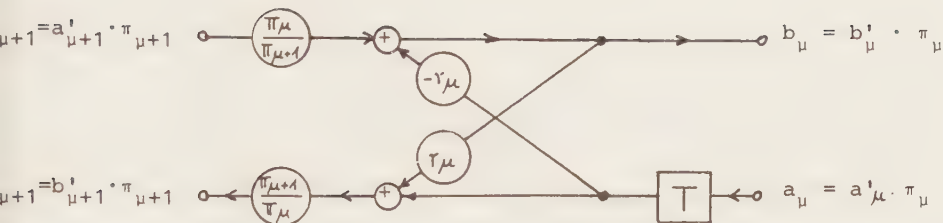


Bild 3.11: Modifizierte Form eines Teil-Zweitores

Entsprechend den veränderten Tor-Signalen $a_{\mu} = a'_{\mu} \pi_{\mu}$, $b_{\mu} = b'_{\mu} \pi_{\mu}$ sind jetzt noch die Ausgangskoeffizienten ξ_{μ} (Bild 3.7) geeignet anzupassen. Definiert man im Fall der Basis-Struktur (Zwei-Multiplizierer-Form) $\xi'_{\mu} := \xi_{\mu}$, so folgt allgemein

$$\xi_{\mu} = \xi'_{\mu} / \pi_{\mu}, \quad \mu = O(1)N \quad (3-79)$$

In Tabelle 3.3 sind mögliche Realisierungsformen der Teil-Zweitore zusammengestellt, die man in der oben beschriebenen Weise herleitet.

Größere Bedeutung kommt der Ein-Multiplizierer-Form und der normierten Form zu. Erstere ist kanonisch bezüglich der Zahl der Speicher und Multiplizierer. Ihr Rundungsgeräusch ist außerdem erheblich besser als das der Zwei-Multiplizierer-Struktur (z.B. /33/). Für $\xi_{\mu} = 0$, $\mu = O(1)n-1$, $\xi_n = 1$, und $\epsilon_{\mu} = +1$, $\mu = O(1)n-1$ ist sie ferner mit einem Wellendigitalfilter identisch, dessen zeitkontinuierliches Referenzfilter aus einer Kaskade von quasireziproken Leitungselementen (QUARL's) besteht /27/. Damit sind auch alle anderen modifizierten Formen in ihrem rekursiven Teil als Wellendigitalfilter zu verstehen. Um dies zu veranschaulichen, betrachten wir die Ein-Multipliziererform als Basis-Struktur. Dann gewinnt man offenbar die übrigen Formen gemäß Bild 3.12. Das vorgeschaltete Zweitor ist aber nichts anderes als die Wellendigitalfilter-Realisierung eines idealen Übertragers mit dem Übersetzungsverhältnis $\bar{u} = \frac{\pi_{\mu+1}}{\pi_{\mu}} (1+r_{\mu})$ (siehe Tabelle 3.2).

BEZEICHNUNG	$\pi_\mu / \pi_{\mu+1}$	α_μ	β_μ	γ_μ	δ_μ	REALISIERUNG
Allgemeine Form	$\pi_\mu / \pi_{\mu+1}$	$\frac{\pi_\mu}{\pi_{\mu+1} + 1}$	$-\pi_\mu$	$[1 - x_\mu^2] \frac{\pi_{\mu+1}}{\pi_\mu}$	x_μ	
Zwei-Multiplizierer-Form	1	1	$-\pi_\mu$	$1 - x_\mu^2$	x_μ	
Ein-Multiplizierer-Form	$1 + \varepsilon_\mu x_\mu$ $\varepsilon_\mu \{+1, -1\}$	$1 + \varepsilon_\mu x_\mu$	$-\pi_\mu$	$1 - \varepsilon_\mu x_\mu$	x_μ	
Normierte Form ("Normalized ladder")	$\sqrt{1 - x_\mu^2}$	$\sqrt{1 - x_\mu^2}$	$-\pi_\mu$	$\sqrt{1 - x_\mu^2}$	x_μ	

Tabelle 3.3: Mögliche Realisierungsformen der Teil-Zweitore

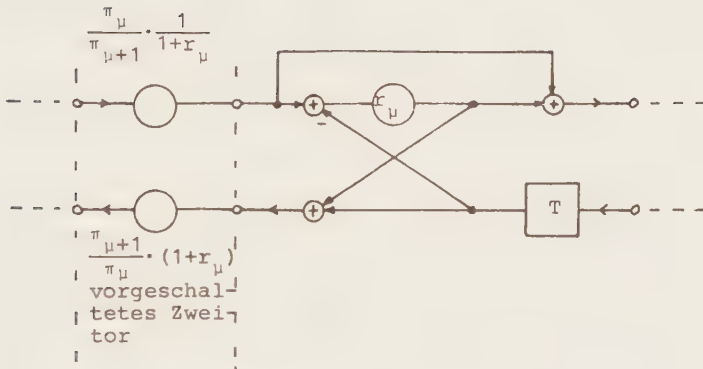


Bild 3.12: Zur Herleitung der modifizierten Strukturen aus der Ein-Multiplizierer-Form mit $\epsilon_\lambda = +1$

3.4.2 Skalierung

Bei der Basis-Struktur (Zwei-Multiplizierer-Form) sind die Ausgänge der internen Tore zunächst unskaliert. Wie in /33/ gezeigt wird, bestehen hier im Innern des Filters beträchtliche Unterschiede bzgl. der Aussteuerung. Die damit verbundene schlechte Ausnutzung des verfügbaren Dynamikbereiches führt zu einem hohen Rundungsgeräusch am Filterausgang. Wie bei den allgemeinen Wellendigitalfiltern ist eine Skalierung durch Einfügen idealer Übertrager möglich, woraus die betrachteten modifizierten Formen resultieren. Hier ist die in Tabelle 3.3 angegebene normierte Form (normalized ladder) von besonderer Bedeutung, bei der sämtliche Torausgänge die L_2 -Norm 1 besitzen /32/. Dementsprechend ist das Rundungsgeräusch deutlich besser als in den anderen Fällen. Im Fall der Ein-Multiplizierer-Form kann eine gleichmäßige Aussteuerung nur näherungsweise durch optimale Wahl der Vorzeichenparameter ϵ_μ erreicht werden /31/.

Durch geeignete Wahl der Koeffizienten π_μ ist im übrigen auch eine Skalierung bzgl. der L_∞ -Norm durchführbar (Diplomarbeit /36/),

eine Möglichkeit, die offenbar bislang in der Literatur nicht berücksichtigt wurde.

3.4.3 Überlaufstabilität

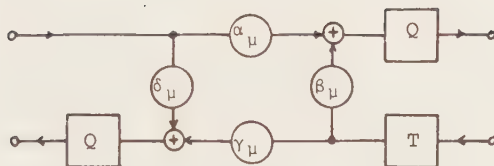
Der Stabilitätsnachweis erfolgt wieder mit Hilfe des in Abschnitt 2.3 eingeführten allgemeinen Filtermodells. Dazu betrachten wir ein Teil-Zweitor in der allgemeinen Form mit Quantisierern an den Tor-Ausgängen (Bild 3.13a). Um den Bezug zu dem genannten allgemeinen Modell herzustellen, wird zunächst die Anordnung von Bild 3.13a in ein 5-Tor gemäß Bild 3.13b umgezeichnet. Die Signale an den Toren wurden hier entsprechend der Nomenklatur von Bild 2.7 bezeichnet. Damit formuliert man den Zusammenhang zwischen allen ein- und auslaufenden Signalen zu

$$\begin{bmatrix} b_1 \\ b_{21} \\ b_{22} \\ b_{31} \\ b_{32} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ \beta_\mu & 0 & 0 & 0 & \alpha_\mu \\ \gamma_\mu & 0 & 0 & 0 & \delta_\mu \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_{21} \\ a_{22} \\ a_{31} \\ a_{32} \end{bmatrix} \quad (3-80)$$

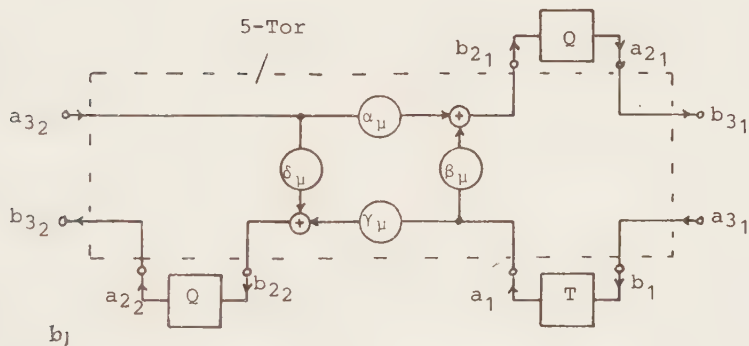
$$\underline{b} = \underline{V} \cdot \underline{a} ,$$

woraus sich entsprechend Gleichung (2-23), Abschnitt 2.3.1, die Verknüpfungsmatrix $\underline{V}$ ergibt. In der allgemeinen Form gilt gemäß (3-78)

$$\alpha_\mu = \frac{\pi_\mu}{\pi_{\mu+1}} , \beta_\mu = -r_\mu , \gamma_\mu = [1-r_\mu^2] \cdot \frac{\pi_{\mu+1}}{\pi_\mu} , \delta_\mu = r_\mu .$$



a)



b)

Bild 3.13:

Damit wird $\underline{V}$ zu

$$\underline{V} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ -r_\mu & 0 & 0 & 0 & \frac{\pi_\mu}{\pi_\mu + 1} \\ [1 - r_\mu^2] \frac{\pi_\mu + 1}{\pi_\mu} & 0 & 0 & 0 & r_\mu \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} =$$

$$= \underline{T} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ -r_\mu & 0 & 0 & 0 & \sqrt{1 - r_\mu^2} \\ \sqrt{1 - r_\mu^2} & 0 & 0 & 0 & r_\mu \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \underline{T}^{-1}$$

$$:= \underline{V}_N$$

(3-81)

$$\underline{T} = \text{Diag} \{1, 1, \frac{\pi_{\mu+1}}{\pi_{\mu}} \sqrt{1-r_{\mu}^2}, 1, \frac{\pi_{\mu+1}}{\pi_{\mu}} \sqrt{1-r_{\mu}^2}\}.$$

Die Matrix $\underline{V}_N$ ist offenbar orthogonal mit $\underline{V}_N^T \underline{V}_N = \underline{E}$. Wir untersuchen nun den Zusammenhang zwischen $||\underline{b}||_{\underline{G}}$ und $||\underline{a}||_{\underline{G}}$, wenn die Vektornormen wie folgt definiert sind:

$$||\underline{a}||_{\underline{G}}^2 = \underline{a}^T \underline{G} \underline{a}, \quad ||\underline{b}||_{\underline{G}}^2 = \underline{b}^T \underline{G} \underline{b} \quad (3-82)$$

mit der positiven Diagonalmatrix

$$\underline{G} = \underline{T}^{-1} \underline{T}^{-1}.$$

Mit (3-80), (3-81) und (3-82) erhält man

$$\begin{aligned} ||\underline{b}||_{\underline{G}}^2 &= \underline{b}^T \underline{T}^{-1} \underline{T}^{-1} \underline{b} = \underline{a}^T \underline{V}^T \underline{T}^{-1} \underline{T}^{-1} \underline{V} \underline{a} = \\ &= \underline{a}^T \underline{T}^{-1} \underbrace{\underline{V}_N^T \underline{T} \underline{T}^{-1} \underline{T}^{-1} \underline{T} \underline{V}_N \underline{T}^{-1}}_{= \underline{E}} \underline{a} = \underline{a}^T \underline{T}^{-1} \underbrace{\underline{V}_N^T \underline{V}_N}_{= \underline{E}} \underline{T}^{-1} \underline{a} = \\ &= \underline{a}^T \underline{T}^{-1} \underline{T}^{-1} \underline{a} = ||\underline{a}||_{\underline{G}}^2. \end{aligned} \quad (3-83)$$

Nach Abschnitt 2.3 ist dies eine hinreichende Bedingung für die inkrementale Pseudopassivität des betrachteten Teil-Zweitores, sofern die Dreiecks- oder Sättigungskennlinie als Überlauf-Charakteristik verwendet wird. Außerdem besteht bei Betragsabschneiden Freiheit von granularen Grenzzyklen.

Für den Spezialfall der normierten Form (normalized ladder) gilt nach Tabelle 3.3

$$\frac{\pi_{\mu+1}}{\pi_{\mu}} = \frac{1}{\sqrt{1-r_{\mu}^2}}$$

so daß $\underline{T}$ zur Einheitsmatrix wird und $\underline{V}$ in die Orthogonalmatrix $\underline{V}_N$ übergeht. Die Operation $\underline{b} = \underline{V} \underline{a}$ bedeutet dann eine Drehung des Vektors $\underline{a}$, wobei die euklidische Norm erhalten bleibt.

Neben dem in Bild 3.13 betrachteten Teil-Zweitor tritt am Ende der Kaskade noch eine einfache Verbindung auf, die als Ein-Tor die Beziehung $b = a$ realisiert. Offenbar ist die Verknüpfungsmatrix eine zum Skalar $V = 1$ entartete Orthogonalmatrix.

Der gesamte rekursive Teil des Filters entsteht durch Kaskadierung obiger Untersysteme und besitzt demnach ebenfalls die Eigenschaft (siehe Abschnitt 2.3.2)

$$||\underline{V} \underline{a}||_{\underline{G}} = ||\underline{a}||_{\underline{G}} \quad \forall \underline{a} \in \mathbb{R}^N \quad (3-84)$$

$\underline{V}$, $\underline{a}$ und $\underline{G}$ kennzeichnen hier das Gesamtsystem.

Bzgl. der Konstruktion von $\underline{G}$ sei nochmals auf Abschnitt 2.3.2 verwiesen, in dessen letztem Teil die Zusammenschaltung von Systemen behandelt wurde.

Im Spezialfall der "normalized ladder"-Struktur wird $\underline{G}$ zur Einheitsmatrix und $\underline{V}$ orthogonal.

3.4.4 Die Systemmatrix A

Aufgrund der allgemeinen Ergebnisse aus Abschnitt 2.3.4 folgt mit

$$||\underline{V} \underline{a}||_{\underline{G}} = ||\underline{a}||_{\underline{G}} \quad \forall \underline{a} \in \mathbb{R}^N$$

die Ungleichung

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n. \quad (3-85)$$

Dabei enthält die Diagonalmatrix $\underline{G}_1$ diejenigen Komponenten von $\underline{G}$, welche zu den Zustandsvariablen gehören. Im Fall der "normalized ladder"-Struktur ist $\underline{G}_1$ die Einheitsmatrix.

Um genauere Aussagen zu erhalten, betrachten wir exemplarisch die $\underline{A}$ -Matrizen für Filtergrade $n = 2, 3, 4$. Dabei wird die all-

gemeine Struktur nach Bild 3.14 zugrunde gelegt.

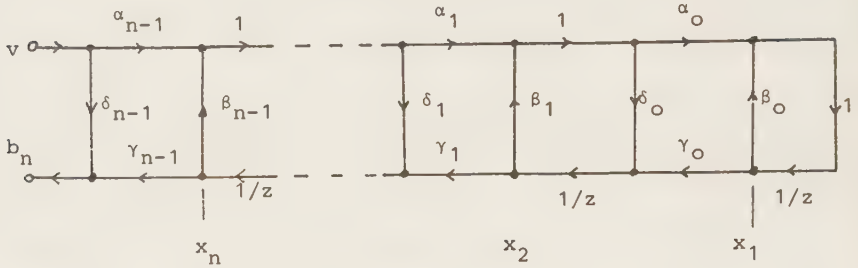


Bild 3.14: Rekursiver Teil der Leiterstruktur in allgemeiner Form

Beispiel 1: $n = 2$

$$\underline{A} = \begin{bmatrix} \beta_0 & \beta_1 & \alpha_0 \\ \gamma_0 & \beta_1 & \delta_0 \end{bmatrix}$$

Mit $\alpha_\mu = \frac{\pi_\mu}{\pi_{\mu+1}}$, $\beta_\mu = -r_\mu$, $\gamma_\mu = [1-r_\mu^2] \frac{\pi_{\mu+1}}{\pi_\mu}$, $\delta_\mu = r_\mu$

(Tabelle 3.3) und geeigneter Umformung ergibt sich

$$\underline{A} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \frac{\pi_1}{\pi_0} \sqrt{1-r_0^2} \end{bmatrix}}_{= \underline{G}_1^{-1/2}} \underbrace{\begin{bmatrix} -r_0 & -r_1 \sqrt{1-r_0^2} \\ \sqrt{1-r_0^2} & -r_1 r_0 \end{bmatrix}}_{:= \underline{\hat{A}}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \frac{\pi_0}{\pi_1} \frac{1}{\sqrt{1-r_0^2}} \end{bmatrix}}_{:= \underline{G}_1^{1/2}} \quad (3-86)$$

$\underline{\hat{A}}$ lässt sich noch in etwas anderer Form schreiben:

$$\underline{\hat{A}} = \begin{bmatrix} -r_0 & -\sqrt{1-r_0^2} \\ \sqrt{1-r_0^2} & -r_0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & r_1 \end{bmatrix} \quad (3-87)$$

Orthogonalmatrix

Beispiel 2: n = 3

$$\underline{A} = \begin{bmatrix} \beta_0 & \beta_1 \alpha_0 & \beta_2 \alpha_1 \alpha_0 \\ \gamma_0 & \beta_1 \delta_0 & \beta_2 \alpha_1 \delta_0 \\ 0 & \gamma_1 & \beta_2 \delta_1 \end{bmatrix}$$

Nach Einsetzen von α_μ , β_μ , γ_μ , δ_μ und Umformung ergibt sich ähnlich Beispiel 1 die Darstellung

$$\underline{A} = \underline{G}_1^{-1/2} \hat{\underline{A}} \underline{G}_1^{1/2}$$

mit

$$\underline{G}_1^{-1/2} = \text{Diag} \left\{ 1, \frac{\pi_1}{\pi_0} \sqrt{1-r_0^2}, \frac{\pi_2}{\pi_0} \sqrt{1-r_0^2} \sqrt{1-r_1^2} \right\}$$

$$\hat{\underline{A}} = \begin{bmatrix} -r_0 & -r_1 \sqrt{1-r_0^2} & -\sqrt{1-r_1^2} \sqrt{1-r_0^2} \\ \sqrt{1-r_0^2} & -r_1 r_0 & r_0 \sqrt{1-r_1^2} \\ 0 & \sqrt{1-r_1^2} & -r_1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & r_2 \end{bmatrix}$$

Orthogonalmatrix

(3-88)

Beispiel 3: n = 4

$$\underline{A} = \underline{G}_1^{-1/2} \hat{\underline{A}} \underline{G}_1^{1/2}$$

mit

$$\underline{G}_1^{-1/2} = \text{Diag} \left\{ 1, \frac{\pi_1}{\pi_0} \sqrt{1-r_0^2}, \frac{\pi_2}{\pi_0} \sqrt{1-r_0^2} \sqrt{1-r_1^2}, \right. \\ \left. \frac{\pi_3}{\pi_0} \sqrt{1-r_0^2} \sqrt{1-r_1^2} \sqrt{1-r_2^2} \right\}$$

$$\hat{\underline{A}} = \underbrace{\begin{bmatrix} -r_0 & -r_1\sqrt{1-r_0^2} & -r_2\sqrt{1-r_0^2}\sqrt{1-r_1^2} & -\sqrt{1-r_0^2}\sqrt{1-r_1^2}\sqrt{1-r_2^2} \\ \sqrt{1-r_0^2} & -r_1r_0 & -r_2r_0\sqrt{1-r_1^2} & -r_0\sqrt{1-r_1^2}\sqrt{1-r_2^2} \\ 0 & \sqrt{1-r_1^2} & -r_2r_1 & -r_1\sqrt{1-r_2^2} \\ 0 & 0 & \sqrt{1-r_2^2} & -r_2 \end{bmatrix}}_{\text{Orthogonalmatrix}}$$

$$\cdot \begin{bmatrix} 1 & 0 & \text{---} & 0 \\ 0 & 1 & \diagdown & 0 \\ & & 1 & 0 \\ 0 & \text{---} & 0 & r_3 \end{bmatrix} \quad (3-89)$$

Offenbar läßt sich in den behandelten Beispielen die $\underline{A}$ -Matrix immer in der Form

$$\underline{A} = \underline{G}_1^{-1/2} \hat{\underline{A}} \underline{G}_1^{1/2} \quad (3-90)$$

darstellen, wobei man $\hat{\underline{A}}$ stets als

$$\hat{\underline{A}} = \underline{U} \cdot \text{Diag} \{1, 1, \dots, 1, r_{n-1}\} \quad (3-91)$$

mit $\underline{U}$ = Orthogonalmatrix

schreiben kann. Es sei vermerkt, daß sich dieses Ergebnis auch für höhergradige Systeme verifizieren läßt. Auf einen allgemeinen Beweis wird jedoch verzichtet.

Im Spezialfall der "normalized ladder" Struktur gilt gemäß Tabelle 3.3

$$\pi_\mu = \pi_{\mu+1} \sqrt{1-r_\mu^2},$$

woraus man

$$\frac{\pi_0}{\pi_v} = \sqrt{1-r_0^2} \sqrt{1-r_1^2} \dots \sqrt{1-r_{v-1}^2}$$

ableitet. Damit wird $\underline{G}_1^{-1/2}$ zur Einheitsmatrix und es folgt $\underline{A} = \hat{\underline{A}}$. Die Operation

$$\underline{A} \underline{x} = \hat{\underline{A}} \underline{x} = \underline{U} \cdot \text{Diag}\{1, \dots, 1, r_{n-1}\} \underline{x}$$

bedeutet demnach eine Abschwächung der n-ten Komponente von $\underline{x}$ um den Faktor r_{n-1} ($|r_{n-1}| < 1$) und anschließende Drehung mittels der Orthogonalmatrix $\underline{U}$. Somit folgt

$$||\underline{A} \underline{x}||_{\underline{E}} \leq ||\underline{x}||_{\underline{E}} \quad \forall \underline{x} \in \mathbb{R}^n. \quad (3-92)$$

Falls die Komponente x_n des Vektors $\underline{x}$ verschwindet, gilt stets das Gleichheitszeichen. Allgemein erhält man

$$||\underline{A} \underline{x}||_{\underline{E}}^2 = ||\underline{x}||_{\underline{E}}^2 - (r - r_{n-1}^2) \cdot x_n^2 \quad (3-93)$$

Dieses Ergebnis läßt sich auf die anderen betrachteten Leiterstrukturen verallgemeinern, wenn man zur Norm-Definition die positive Diagonalmatrix $\underline{G}_1$ heranzieht. (3-93) geht dann über in

$$||\underline{A} \underline{x}||_{\underline{G}_1}^2 = ||\underline{x}||_{\underline{G}_1}^2 - g_n [1 - r_{n-1}^2] x_n^2, \quad (3-94)$$

wobei g_n das letzte Hauptdiagonalelement von $\underline{G}_1$ bezeichnet.

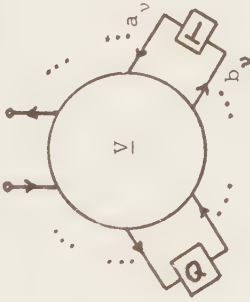
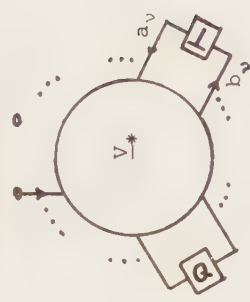
4. ZUSAMMENFASSUNG UND ERGÄNZUNG DER BISHERIGEN ERGEBNISSE

4.1 Modelle zur Analyse der Überlaufstabilität

In Kapitel 2 wurden hinreichende Bedingungen für die Überlaufstabilität von Digitalfiltern angegeben, wobei unterschiedliche Beschreibungsformen der Struktur zur Anwendung kamen. Zusammen mit speziellen einfachen Überlaufcharakteristiken (z.B. Dreiecks- oder Sättigungskennlinie) gewährleisteten die genannten Bedingungen die Freiheit von Überlaufgrenzzyklen sowie die Antwortstabilität bei Erregung. Außerdem sind bei verschwindender Erregung und Betragsabschneiden keine granularen Grenzzyklen möglich.

Die theoretischen Beschreibungsformen und ihre Anwendungsbereiche sind noch einmal in Tabelle 4.1 zusammengestellt. Neben der bekannten Analyse anhand der Systemmatrix $\underline{A}$ (Zustandsraum-Darstellung) wird ein Allgemeines Filtermodell zugrundegelegt, welches auf beliebige Strukturen angewandt werden kann. Es gestattet insbesondere auch einen einfachen Nachweis der Überlaufstabilität von Wellendigitalfiltern, der bislang mit der "Pseudopassivität" der einzelnen Bausteine erbracht wurde. Es existieren zwei Varianten des allgemeinen Modells. Variante I beschreibt das Gesamtsystem einschließlich der unbeschalteten Ein- und Ausgänge. Wie gezeigt wurde, bleibt beim Zusammenschalten von Systemen die Stabilität erhalten, sofern die Teilfilter diese Eigenschaft besitzen. Diese Tatsache läßt sich z.B. vorteilhaft bei den Leiterstrukturen (Markel / Gray) nützen, welche in ihrem rekursiven Teil aus überschaubaren gleichartigen Teilsystemen aufgebaut sind.

Die für diese erste Variante angegebene Stabilitätsbedingung kennzeichnet inkremental pseudopassive Filter. Sie ist allerdings in gewissen Fällen zu streng. Dies führt zur Variante II des allgemeinen Modells. Hierbei werden die zu den unbeschalteten Ausgängen führenden Signalfade gestrichen. Die linearen arithmetischen Operationen werden nun durch die modifizierte Verknüpfungs-

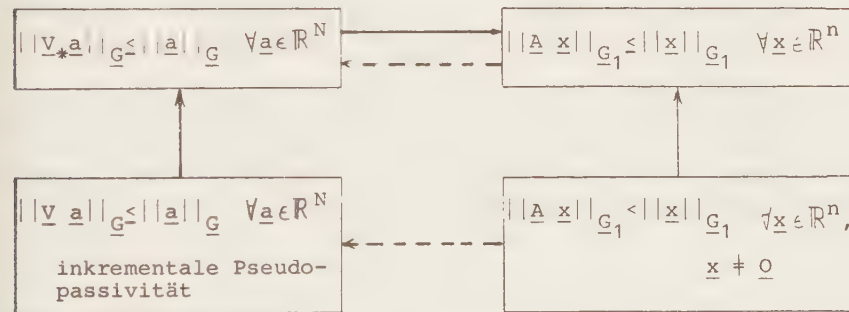
Anwendungsbereich	Strukturen mit Quantisieren (Überlaufkennlinien) <u>nur</u> an den Eingängen der Zustandspeicher	Strukturen mit Quantisieren an beliebigen Stellen im Filter (Spezialfälle: Strukturen mit Quantisieren am Eingang der Zustandspeicher)
Beschreibungsform	Zustandsraum-Darstellung $\underline{x}(k+1) = \underline{A} \underline{x}(k) + \underline{B} \underline{v}(k)$	Allgemeines Filtermodell I. Gesamt-System mit Ein-/Ausgängen  II. Rekursiver Teil + Eingänge, Ausgänge gleich Null gesetzt  $\underline{b}(k) = \underline{V} \underline{a}(k)$ $\underline{b}(k) = \underline{V}_* \underline{a}(k)$
hinreichende Bedingung für Überlaufstabilität	$\ \underline{A} \underline{x}\ _{\underline{G}_1} \leq \ \underline{x}\ _{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n$	$\ \underline{b}\ _{\underline{G}} = \ \underline{V} \underline{a}\ _{\underline{G}} \leq \ \underline{a}\ _{\underline{G}} \quad \forall \underline{a} \in \mathbb{R}^N$ → inkrementale Pseudopassivität

Anwendungsbereich	Strukturen mit Quantisieren (Überlaufkennlinien) nur an den Eingängen der Zustandspeicher	Strukturen mit Quantisieren an beliebigen Stellen im Filter (Spezialfälle: Strukturen mit Quantisieren am Eingang der Zustandspeicher)
Beispiele für die Anwendung	- Zustandsraum-Strukturen . Filter nach Barnes ($\underline{A}$ normal) . Rauschoptimaler Block zweiten Grades nach Mullis/Roberts - Block zweiten Grades in 1./2. kanonischer Form	- Wellendigitalfilter - Leiterstrukturen nach Markel / Gray

Tabelle 4.1: Beschreibungsformen zur Analyse der Überlaufstabilität und ihre Anwendung

matrix $\underline{V}_*$ beschrieben, welche aus $\underline{V}$ hervorgeht, wenn man die zu den genannten Ausgängen gehörenden Zeilen gleich Null setzt. Die Stabilitätsbedingung lautet jetzt $||\underline{V}_* \underline{a}||_{\underline{G}} \leq ||\underline{a}||_{\underline{G}}$. Sie gewährleistet zwar nicht mehr die inkrementale Pseudopassivität im strengen Sinn, doch beziehen sich die Unterschiede nur auf die unbeschalteten Ausgänge. Das Stabilitätsverhalten des rekursiven Teils entspricht dem inkremental pseudopassiver Systeme. Die zweite Variante gestattet i.a. jedoch keine Aussagen bzgl. des Zusammenschaltens von Teilfiltern.

Die folgende schematische Darstellung zeigt die Zusammenhänge zwischen den Eigenschaften der Verknüpfungsmatrizen $\underline{V}$ und $\underline{V}_*$ sowie der Systemmatrix $\underline{A}$.



Der Schluß von rechts nach links gilt jeweils nur dann, wenn die Quantisierer am Eingang der Zustandsspeicher auftreten (gestrichelte Pfeile).

4.2 Eigenschaften der Systemmatrix A, Vergleich der Filterstrukturen

Alle in Kapitel 3 beschriebenen Strukturen zeichnen sich durch die gemeinsame Eigenschaft der Antwortstabilität aus. In allen Fällen gilt

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n \quad (2-11)$$

$$\text{bzw. } \sqrt{(\underline{A} \underline{x})^T \underline{G}_1 (\underline{A} \underline{x})} \leq \sqrt{\underline{x}^T \underline{G}_1 \underline{x}}.$$

Ein möglicher Vergleich kann sich z.B. auf die Wahl der Diagonalmatrix $\underline{G}_1$ beziehen. Hier weisen die Normalform-Filter ($\underline{A}$ ist normal) und die "normalized ladder" Struktur nach Markel und Gray Parallelen auf. In beiden Fällen kann $\underline{G}_1 = \underline{E}$ gesetzt werden, eine Eigenschaft, die insbesondere für die Stabilität zeitvariabler Filter von Interesse ist (siehe Kap. 7).

Für die folgenden Betrachtungen soll anstelle von $\underline{A}$ die Matrix

$$\hat{\underline{A}} = \underline{G}_1^{1/2} \underline{A} \underline{G}_1^{-1/2} \quad (4-1)$$

herangezogen werden. Damit läßt sich (2-11) in der Form

$$\|\hat{\underline{A}}(\underline{G}_1^{1/2} \underline{x})\|_{\underline{E}} \leq \|\underline{G}_1^{1/2} \underline{x}\|_{\underline{E}} \quad \forall \underline{G}_1^{1/2} \underline{x} \in \mathbb{R}^n \quad (4-2)$$

schreiben, wobei nun immer die euklidische Vektornorm auftritt. Die Ähnlichkeitstransformation (4-1) mit der Diagonalmatrix $\underline{G}_1^{1/2}$ ist im übrigen als einfache Umskalierung des Filters zu interpretieren. Der neue ZV-Vektor ist dann $\hat{\underline{x}} = \underline{G}_1^{1/2} \underline{x}$ und mit (4-2) gilt

$$\|\hat{\underline{A}} \hat{\underline{x}}\|_{\underline{E}} \leq \|\hat{\underline{x}}\|_{\underline{E}} \quad \forall \hat{\underline{x}} \in \mathbb{R}^n. \quad (4-3)$$

Wie sich gezeigt hat, ist $\hat{\underline{A}}$ in vielen Fällen eine normale Matrix. Hier ist die Supremumsnorm

$$\max_{\substack{\hat{\underline{x}} \in \mathbb{R}^n \\ \hat{\underline{x}} \neq \underline{0}}} \frac{\|\hat{\underline{A}} \hat{\underline{x}}\|_{\underline{E}}}{\|\hat{\underline{x}}\|_{\underline{E}}} = \max_v \{|\lambda_v|\}, \quad (4-4)$$

λ_v : Eigenwerte von $\hat{\underline{A}}$ bzw. $\underline{A}$

minimal. Der andere Extremfall liegt vor, wenn die Supremumsnorm gleich 1 wird. Als Beispiel seien die von beidseitig ohm'sch abgeschlossenen LC-Abzweigstrukturen abgeleiteten Wellendigitalfilter genannt; ihre $\hat{\underline{A}}$ -Matrix ist weder normal noch kann sie durch Ähn-

lichkeitstransformation mit einer Diagonalmatrix normal gemacht werden.

Eine Gegenüberstellung der untersuchten Filterstrukturen findet sich in Tabelle 4.2.

Zwei Besonderheiten der Matrix $\hat{\underline{A}}$ sollen noch einmal genauer beleuchtet werden:

1. Im Fall der Wellendigitalfilter hat sich gezeigt, daß $\hat{\underline{A}}$ immer als Untermatrix einer Orthogonalmatrix interpretiert werden kann.
2. Bei den Leiterstrukturen nach Markel und Gray ist $\hat{\underline{A}}$ das Produkt einer Orthogonalmatrix und einer Normalmatrix mit reellen Eigenwerten $\tilde{\lambda}_v$, $|\tilde{\lambda}_v| \leq 1$.

Diese beiden Eigenschaften gelten jedoch weit allgemeiner für überlaufstabile Strukturen. Erstere ist äquivalent zur Stabilitätsbedingung (2-11) bzw. (4-3). Die gleiche Äquivalenz gilt auch im zweiten Fall, sofern alle Eigenwerte von $\hat{\underline{A}}$ ungleich Null sind. Der Beweis erfolgt in Anhang A6.

Unter Berücksichtigung der im vorangegangenen Abschnitt 4.1 angesprochenen Stabilitätseigenschaften lassen sich die Filter schließlich in einem Klassifikationsschema gemäß Bild 4.1 anordnen. Die mit ausgezogenen Linien umrandeten Gebiete kennzeichnen allgemeine Strukturen, bei denen die Quantisierer an beliebigen Stellen auftreten können. Ihre Lage wird durch die Verknüpfungsmatrizen $\underline{V}_*$ bzw. $\underline{V}$ beschrieben. Innerhalb der gestrichelt umrandeten Teile sind dagegen Quantisierer nur an den Eingängen der Zustandsspeicher zugelassen, wegen dieser Besonderheit wird hier das Stabilitätsverhalten nicht nur durch $\underline{V}_*$ bzw. $\underline{V}$ sondern bereits durch die Eigenschaften von $\underline{A}$ vollständig beschrieben. Neben einer Reihe von Strukturen, die generell überlaufstabil sind (Vor: Stabilität im linearen Fall), ist zusätzlich die zweite kanonische Form (Grad 2) eingezeichnet; die Stabilitätsbedingung ist hier nur für einen eingeschränkten Bereich der Polstellenlagen erfüllt.

Struktur	Eigenschaften von $\hat{A} = \underline{G}_1^{-1/2} A \underline{G}_1^{-1/2}$	Supremumsnorm $\frac{ \hat{A} \hat{x} _{\underline{E}}}{ \hat{x} _{\underline{E}}} = \max_{\hat{x} \neq 0} \frac{ \hat{A} \hat{x} _{\underline{E}}}{ \hat{x} _{\underline{E}}} = \max_{\underline{x} \neq 0} \frac{ A \underline{x} _{\underline{G}_1}}{ \underline{x} _{\underline{G}_1}}$	Diagonalmatrix $\underline{G}_1$
Normalform n beliebig	$A = \hat{A}$ normal speziell: $n=2, \lambda_{1,2} = \rho e^{\pm j\psi}$ $\hat{A} = \hat{A} = \rho \begin{bmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{bmatrix}$	allg.: $\max_v \{ \lambda_v \}$ + ρ	$\underline{G}_1 = \underline{E}$
Zustandsraum-Strukturen rauschopt. Form (Mul- lis, Ro- berts) $n = 2$	$\hat{A}$ normal speziell: $n=2, \lambda_{1,2} = \rho e^{\pm j\psi}$ $\hat{A} = \rho \begin{bmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{bmatrix}$	$\max \{ \lambda_{1,2} \}$ + ρ	allg.: $\underline{G}_1 \neq \underline{E}$ für $ \lambda_{1,2} \neq 1$: $\underline{G}_1 \neq \underline{E}$ für spezielle Pol-/ Nullstellen-Konfigu- rationen ($\lambda_{1,2}$ reell): $\underline{G}_1 = \underline{E}$

R, L, Ü oder R, C, Ü-Referenzfilter	$\underline{A}$ symmetrisch $\rightarrow$ normal λ_v reell, $v = 1(1)n$	$\max_v \{ \lambda_v \}$	Elemente von $\underline{G}_1$: Torleitwerte d. Induktivitäten bzw. Kapazitäten d. Referenzfilters
R, C, L, Ü-Referenzfilter	$\hat{\underline{A}} = \begin{bmatrix} -\underline{E}_{n_1} & \underline{E}_{n_2} \end{bmatrix} \hat{\underline{A}}$ $\underline{E}_v = (n_v \times n_v) \text{ Einheitsmatrix}$ $n_1 = \text{Zahl d. Induktiv.}$ $n_2 = \text{Zahl d. Kapazit.}$ $\underline{A}$ symmetrisch $\rightarrow$ normal aber: $\hat{\underline{A}}$ i. a. nicht mehr normal	1 wenn Zahl der Reaktanzen > Zahl der Widerstände (hinreichende Bedingung) z.B. beidseitig ohm'sch abgeschlossene L, C-Abzweigstruktur	Elemente von $\underline{G}_1$: Torleitwerte d. Reaktanzen
R, L, C, Ü, Zirkulatoren, Gyrationen	$\hat{\underline{A}}$ ist Untermatrix einer Orthogonalmatrix	≤ 1 genauere Angabe nur im konkreten Einzelfall möglich	Elemente von $\underline{G}_1$: Torleitwerte der Reaktanzen
Leiterstrukturen (Markel/Gray)	$\hat{\underline{A}} = \underline{U} \cdot \text{Diag}\{1, 1, \dots, 1, r_{n-1}\}$ Normalmatrix Orthogonalmatrix nicht normal	1	$\underline{G}_1^{-1/2} = \text{Diag}\left\{1, \frac{1}{\pi_0} \sqrt{1-r_0^2}, \frac{\pi_2}{\pi_0} \sqrt{1-r_0^2} \sqrt{1-r_1^2}, \dots\right\}$ $\frac{\pi_3}{\pi_0} \sqrt{1-r_0^2} \sqrt{1-r_1^2} \sqrt{1-r_2^2}, \dots$ = $\underline{E}$ für "normalized ladder"

Tabelle 4.2: Vergleich der Filterstrukturen bzgl. ihrer $\underline{A}$ -Matrix

Abschließend soll noch ein interessanter Nebenaspekt angesprochen werden, der sich insbesondere auf $\underline{A}$ -Matrizen bezieht, welche durch Transformation mit einer Diagonalmatrix normal gemacht werden können:

Dazu betrachten wir zunächst eine Lyapunov-Funktion $L(k)$ für beliebige, im linearen Fall stabile Digitalfilter und definieren

$$L(k) := \underline{x}^T(k) \underline{M}^{-*T} \underline{M}^{-1} \underline{x}(k) = \underline{x}^{*T}(k) \underline{M}^{-*T} \underline{M}^{-1} \underline{x}(k) \quad (4-6)$$

$\underline{M}$ = Modalmatrix von $\underline{A}$

$$\text{mit } \underline{A} = \underline{M} \underline{\Lambda} \underline{M}^{-1}$$

$\underline{\Lambda}$ = Diagonalmatrix der Eigenwerte λ_v

Dann gilt bei verschwindender Erregung

$$\begin{aligned} L(k+1) &= [\underline{A} \underline{x}(k)]^T \underline{M}^{-*T} \underline{M}^{-1} [\underline{A} \underline{x}(k)] \\ &= [\underline{A} \underline{x}(k)]^{*T} \underline{M}^{-*T} \underline{M}^{-1} [\underline{A} \underline{x}(k)] \\ &= \underline{x}^{*T}(k) \underline{M}^{-*T} \underline{\Lambda}^* \underline{M}^{*T} \underline{M}^{-*T} \underline{M}^{-1} \underline{M} \underline{\Lambda} \underline{M}^{-1} \underline{x}(k) \\ &= \underline{x}^{*T}(k) \underline{M}^{-*T} \underline{\Lambda}^* \underline{\Lambda} \underline{M}^{-1} \underline{x}(k) \\ &\leq \max_v \{ |\lambda_v|^2 \} \underline{x}^{*T}(k) \underline{M}^{-*T} \underline{M}^{-1} \underline{x}(k) = \max_v \{ |\lambda_v|^2 \} \cdot L(k) . \end{aligned} \quad (4-7)$$

Ist nun $\hat{\underline{A}}$ normal mit

$$\underline{\hat{A}} = \underline{G}_1^{-1/2} \underline{\hat{A}} \underline{G}_1^{1/2} \quad (4-8)$$

und $\underline{\hat{A}} = \underline{U} \underline{\Lambda} \underline{U}^{*T}$ ($\underline{U}$ unitär mit $\underline{U}^{-1} = \underline{U}^{*T}$),

so erhält man die Modalmatrix

$$\underline{M} = \underline{G}_1^{-1/2} \underline{U} \quad \text{bzw.} \quad \underline{M}^{-1} = \underline{U}^{*T} \underline{G}_1^{1/2}$$

und es folgt

$$\underline{M}^{-*T} \underline{M}^{-1} = \underline{G}_1^{1/2} \underline{U} \underline{U}^{*T} \underline{G}_1^{1/2} = \underline{G}_1 . \quad (4-9)$$

Gemäß (4-6) ergibt sich die Lyapunov-Funktion

$$L(k) = \underline{x}^T(k) \underline{G}_1 \underline{x}(k) = ||\underline{x}(k)||_{\underline{G}_1}^2 . \quad (4-10)$$

Sie ist offenbar mit dem Quadrat der bislang verwendeten Norm des ZV-Vektors identisch.

Als Nebenergebnis gilt folgende Aussage:

Jede Matrix $\underline{A}$ mit einfachen Eigenwerten, für die das Produkt $\underline{M}^{-*T} \underline{M}^{-1}$ eine Diagonalmatrix $\underline{G}_1$ mit positiven Elementen ist, erfüllt die Bedingung

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq \max_v \{ |\lambda_v| \} ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n .$$

5. S T R U K T U R V E R G L E I C H B E Z Ü G L I C H D E S R U N D U N G S R A U S C H E N S U N D D E R G R A N U L A R E N G R E N Z Z Y K L E N

5.1 Vorbemerkungen

Bei der Realisierung digitaler Filter in Festkomma-Arithmetik ist u.a. die Varianz des Rundungsrauschens am Ausgang von wesentlicher Bedeutung. Hierdurch wird z.B. bei vorgegebener max. Störleistung die innere Wortlänge des Systems festgelegt.

Wir wollen uns auf den Fall der echten Rundung beschränken. Die Störleistung ist hier geringer als im Fall des Betragsabschneidens, es treten insbesondere i.a. keine Korrelationen zwischen Fehler und Signal auf. Zwar wären die betrachteten Strukturen bei Betragsabschneiden frei von granularen Grenzyklen, doch gilt dies nur bei Erregung mit $v(k) \equiv 0$. Bei konstantem oder periodischem von Null verschiedenem Eingangssignal ist dennoch mit Störszillationen zu rechnen.

Die Überlaufstabilität bleibt auch bei Rundung erhalten.

Zur Skalierung der Systeme wird die L_∞ -Norm herangezogen. Sie ist erfahrungsgemäß ein besseres Kriterium zur Beurteilung der Aussteuerung als etwa die L_2 -Norm. Im Fall der Zustandsraum-Strukturen nach Mullis/Roberts und Barnes, deren Entwurf zunächst auf der Grundlage der L_2 -Norm erfolgt, ist dann allerdings eine nachträgliche Umskalierung notwendig, was durch Transformation des Zustandsvariablen-Vektors mit einer geeigneten Diagonalmatrix erreicht wird.

5.2 Untersuchte Strukturen

Folgende Strukturen, von denen die ersten drei bereits in Kap. 3 beschrieben wurden, sollen untersucht werden:

1. Rauschoptimale Zustandsraum-Struktur nach Mullis und Roberts:

- a) Blöcke zweiten Grades
- b) Parallelanordnung von Blöcken zweiten Grades (die Blockstrukturierung ist wegen der geforderten Überlaufstabilität notwendig).

2. Normalform-Struktur nach Barnes

- a) Blöcke zweiten Grades
- b) Höhergradige Filter, die nicht blockstrukturiert sind.

Anmerkungen:

- Wegen der nachträglichen L_{∞} -Skalierung wird die A-Matrix i.a. nicht mehr normal sein, die Stabilitätseigenschaften bleiben jedoch erhalten.
- Bei Blöcken zweiten Grades ist zu erwarten, und wird sich auch bestätigen, daß die Geräuschleistung größer ist als im Fall der rauschoptimalen Form nach Mullis und Roberts. Bei höhergradigen Filtern ist dagegen ein günstigeres Abschneiden der Struktur nach Barnes denkbar, da hier auf eine Blockstrukturierung verzichtet werden kann und daher zusätzliche Freiheitsgrade für die Minimierung des Rundungsrauschens bestehen. Der Realisierungsaufwand wird sich gegenüber einer Parallel- oder Kaskadenanordnung nicht wesentlich erhöhen, wenn eine segmentweise Verarbeitung mit dem in Kap. 6 beschriebenen Algorithmus erfolgt.

3. Leiterstruktur nach Markel und Gray

Hier sollen insbesondere die Ergebnisse für die allgemeine 4-Multiplizier-Form angegeben werden, da nur diese Struktur die gewünschte L_∞ -Skalierung aller interessierender Punkte im Innern (Ausgänge der Teil-Zweitore) zuläßt. Das Rundungsgeräusch ist hier wesentlich geringer als im Fall der 1- oder 2-Multiplizierer-Form. Ein Vergleich wurde z.B. in /36/ durchgeführt. Dort wurden allerdings die Besonderheiten einer überlaufstabilen Realisierung der 1- und 2-Multipliziererform nicht berücksichtigt, weshalb auf eine Wiedergabe der diese beiden Strukturen betreffenden Ergebnisse verzichtet wird.

4. Struktur nach Meerkötter und Wegener

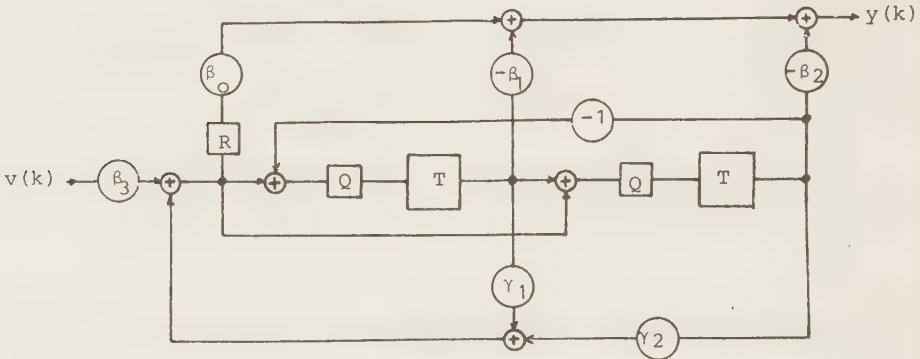
In /29/ wurden von Meerkötter und Wegener zwei Filterblöcke zweiten Grades vorgeschlagen, die in ihrem rekursiven Teil der Wellendigitalfilter-Realisierung eines R,L,C-Schwingkreises entsprechen. Die so gewonnenen Strukturen verbinden somit den Vorteil der Überlaufstabilität von Wellendigitalfiltern mit der Möglichkeit eines einfachen modularen Aufbaus höhergradiger Systeme (z.B. Parallelform).

Die beiden genannten Strukturen sind bzgl. ihres Realisierungsaufwandes gleichwertig. Bezüglich des Rundungsrauschens und der granularen Grenzyklen erweist sich jedoch die in Bild 5.1 angegebene zweite Form als deutlich überlegen /39/. Auf eine Betrachtung der ungünstigeren ersten Struktur wird daher im folgenden verzichtet.

Es sei vermerkt, daß in der Originalarbeit /29/ die Koeffizienten β_0 und β_3 gleich 1 gesetzt wurden. Die modifizierte Form mit $\beta_0 \neq 1$, $\beta_3 \neq 1$ gestattet die Skalierung des Blockausganges sowie einer der beiden Zustandsvariablen. β_3 wird insbesondere so gewählt, daß die L_∞ -Norm der Zustandsvariablen mit der jeweils größeren Aussteuerung gleich 1 wird.

Bei höhergradigen Systemen, die als Parallelanordnung von Blöcken zweiten Grades realisiert sind, läßt sich ein Multiplizierer ein-

sparen, wenn man den Einfluß aller Koeffizienten β_0 in einem direkten Signalpfad zwischen Ein- und Ausgang zusammenfaßt.



R Rundungsstelle

Q Rundung + Überlaufkennlinie

$$\text{Übertragungsfunktion } H(z) = \beta_3 \frac{\beta_0 z^2 - (\beta_1 + \beta_2)z + (\beta_0 + \beta_1 - \beta_2)}{z^2 - (\gamma_1 + \gamma_2)z + (1 + \gamma_1 - \gamma_2)}$$

Bild 5.1: 2. Struktur nach Meerkötter und Wegener

5. Struktur mit Blöcken zweiten Grades in erster kan. Form

Das Verhalten dieser Struktur bezüglich Rundungsgeräusch und Grenzzyklen ist allgemein bekannt (z.B. /5/, /51/). Sie wird daher als Vergleichsmaßstab für die anderen behandelten Realisierungen herangezogen.

5.3 Rundungsrauschen

Die Untersuchungen orientieren sich an den von Dehner /5/ durchgeführten Tests. Somit ist, zumindest teilweise, ein Vergleich mit den dort behandelten Strukturen möglich.

Die Skalierung der Zustandsvariablen sowie des Filterausganges erfolgt gemäß der L_∞ -Norm.

a) Pol-Nullstellen-Test

Hier wird ein Filterblock zweiten Grades mit Nullstellen auf dem Einheitskreis $z_{01,2} = e^{\pm j\pi/6}$ und einem konjugiert komplexen Polstellenpaar $z_{\infty 1,2} = \rho e^{\pm j\psi}$ betrachtet. Nun wird bei festgehaltener Nullstelle und vorgegebenem Polradius ρ der Polwinkel ψ in den Grenzen $0 < \psi < \pi$ variiert (Bild 5.2) und jeweils die innere Rauschkennzahl

$$R_1 = \frac{\sigma^2}{Q^2/12} \quad (5-1)$$

σ^2 = Varianz des Rundungsrauschens am Filterausgang

Q = Quantisierungsintervall

ermittelt.

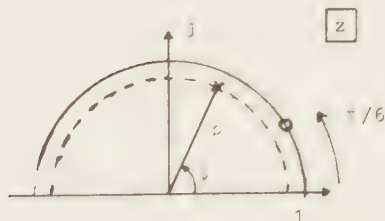


Bild 5.2: Pol-Nullstellen-Test

II. Test eines höhergradigen Filters

Als Prototyp-Filter wird ein aus /5/, Beispiel 4, übernommener Cauer-Tiefpaß 10-ten Grades herangezogen. Seine Übertragungsfunktion lautet

$$H(z) = 0,00695 \prod_{v=1}^5 \frac{z^2 + b_{1v}z + 1}{z^2 + c_{1v}z + c_{0v}} = 0,00695 + \sum_{v=1}^5 \frac{d_{1v}z + d_{0v}}{z^2 + c_{1v}z + c_{0v}}$$

Die zugehörigen Koeffizienten b_{1v} , c_{1v} und c_{0v} finden sich in Tabelle 5.1

v	b_{1v}	c_{1v}	c_{0v}	d_{1v}	d_{0v}
1	-0,751924	-0,913424	0,952110	-0,047888	0,055450
2	-0,657665	-0,889780	0,836577	0,213035	0,003890
3	-0,388098	-0,873145	0,659143	0,118590	-0,488646
4	0,299157	-0,857794	0,415948	-1,240382	0,644686
5	1,636252	-0,846440	0,210392	0,988064	-0,152973

Tabelle 5.1: Koeffizienten des Cauer-Tiefpasses 10-ten Grades

Ferner gelten folgende Spezifikationen:

Normierte Durchlaß-Grenzfrequenz $\Omega_D = \pi/3 = 1,0472$

Normierte Sperr-Grenzfrequenz $\Omega_S = 1,18$

Toleranz im Durchlaßbereich $\delta_D = 0,02$

Toleranz im Sperrbereich $\delta_S = 0,001 \hat{=} -60 \text{ dB}$

Die Variation der Grenzfrequenz erfolgt mittels der Transformation (5-2), wobei folgender Zusammenhang zwischen α und den Grenzfrequenzen Ω_D und Ω'_D vom Prototyp und transformiertem Tiefpaß besteht:

$$\alpha = \frac{\tan \frac{\Omega_D'}{2} - \tan \frac{\Omega_D}{2}}{\tan \frac{\Omega_D'}{2} + \tan \frac{\Omega_D}{2}}$$

Die innere Rauschkennzahl R_i wird in Abhängigkeit von Ω_D' ermittelt.

Die Ergebnisse der durchgeführten Tests sind in den Bildern 5.4 bis 5.13 zusammengestellt.

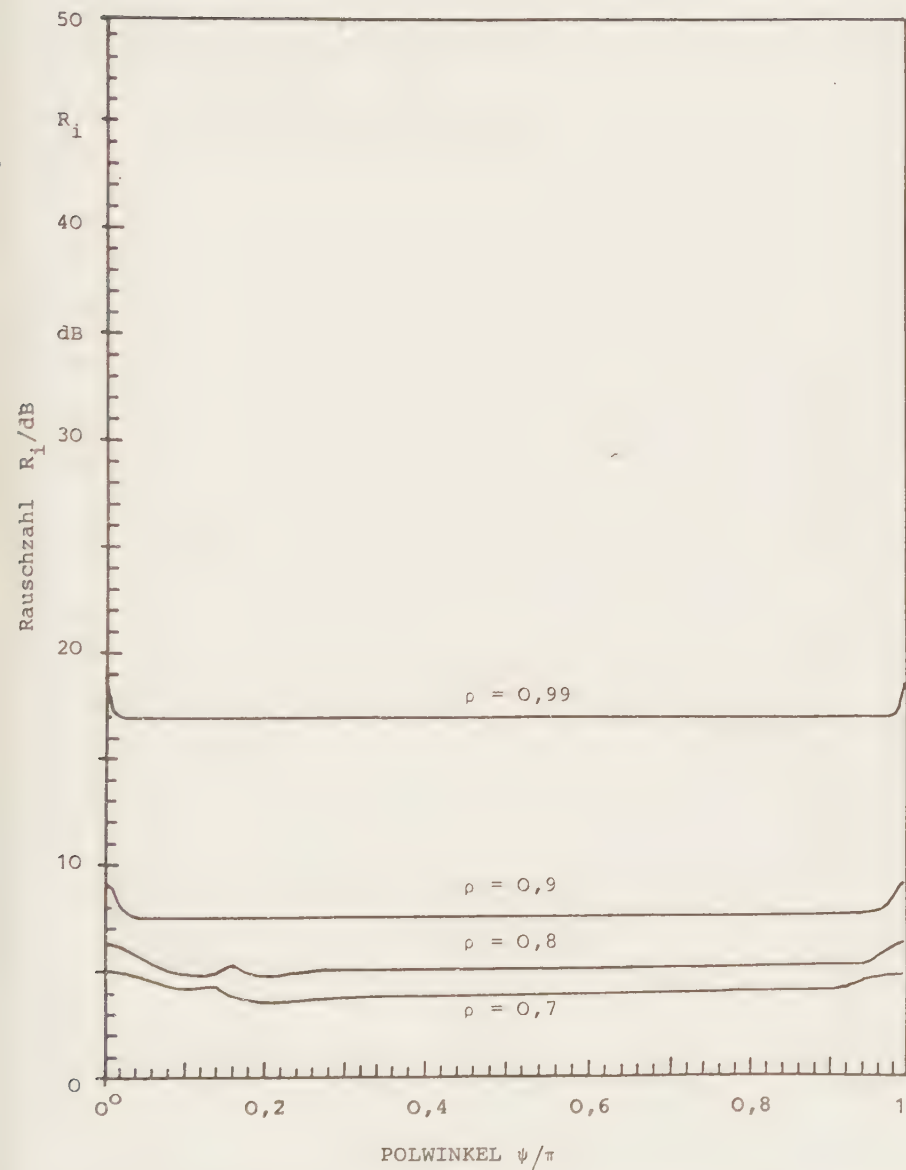


Bild 5.4: Pol-Nullstellen-Test für einen Block zweiten Grades nach Mullis und Roberts

Nullstellen: $z_{0,2} = e^{\pm j\pi/6}$

Polstellen: $z_{\infty,2} = \rho e^{\pm j\psi}$

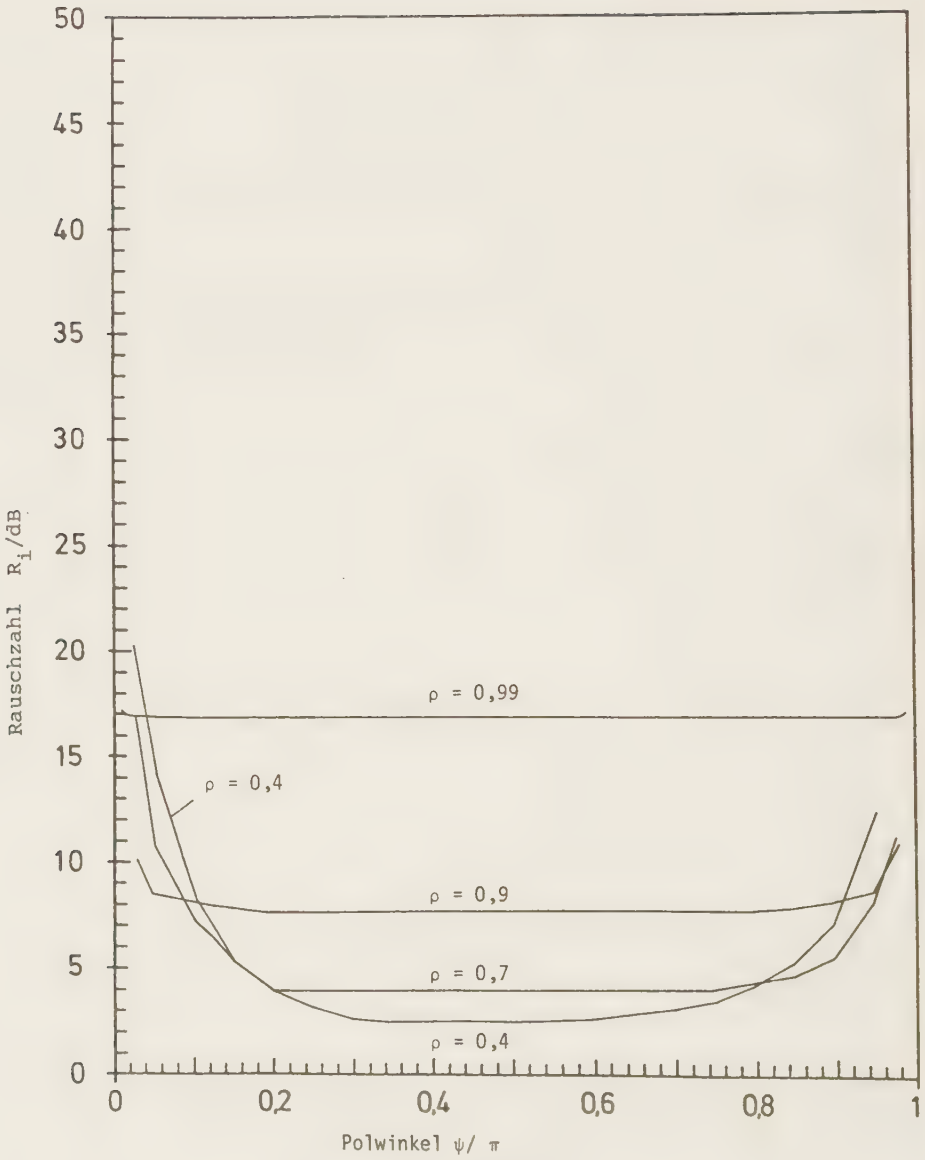


Bild 5.5: Pol-Nullstellen-Test der Struktur nach Barnes

Nullstellen: $z_{01,2} = e^{\pm j\pi/6}$

Polstellen: $z_{1,2} = \rho e^{\pm j\psi}$

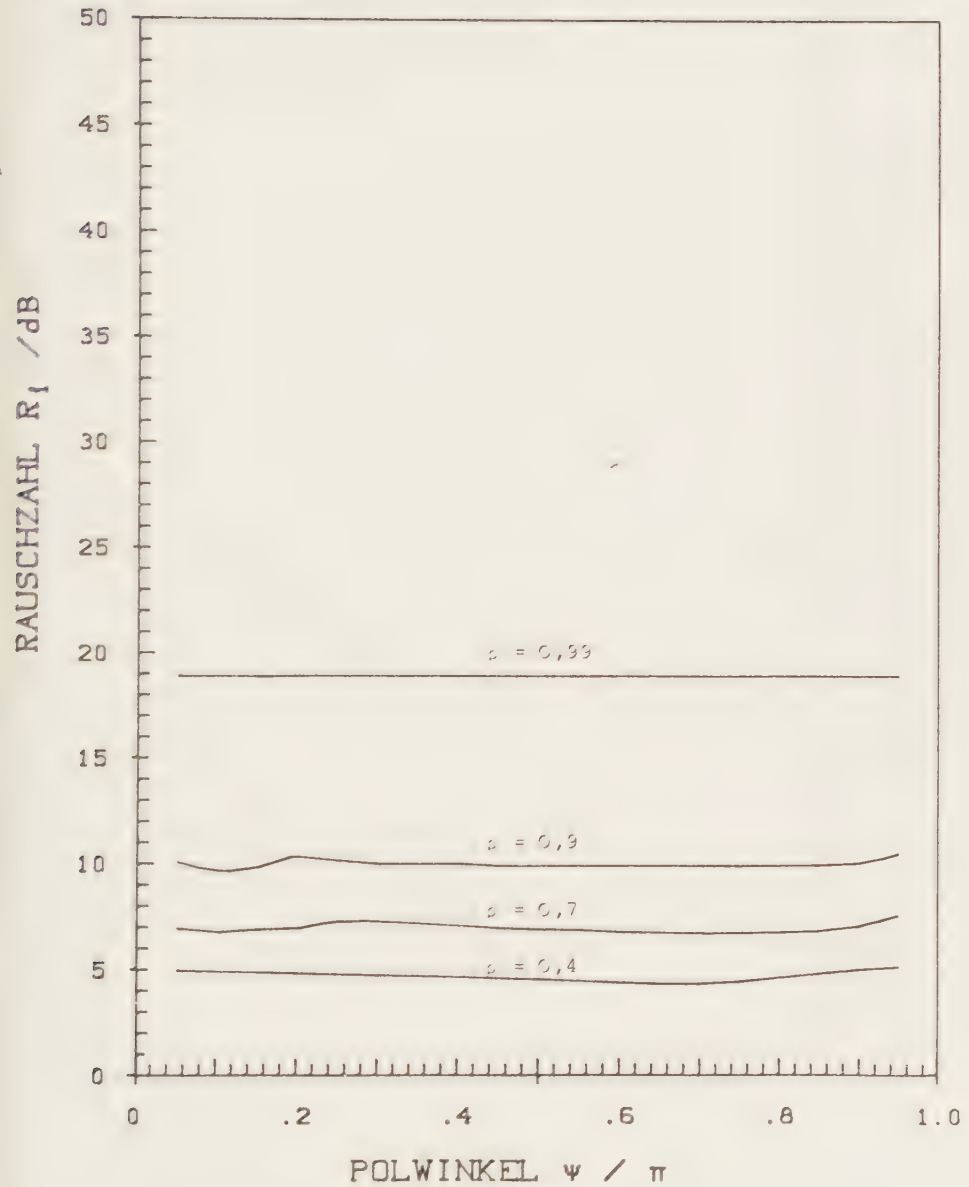


Bild 5.6: Pol-Nullstellen-Test der Leiterstruktur nach Markel und Gray (4-Multiplizierer-Form) mit L_∞ -Skalierung an den Ausgängen der Teil-Zweitore

Nullstelle: $z_{01,2} = e^{\pm j\pi/6}$

Polstelle: $z_{\infty 1,2} = \rho e^{\pm j\varphi}$

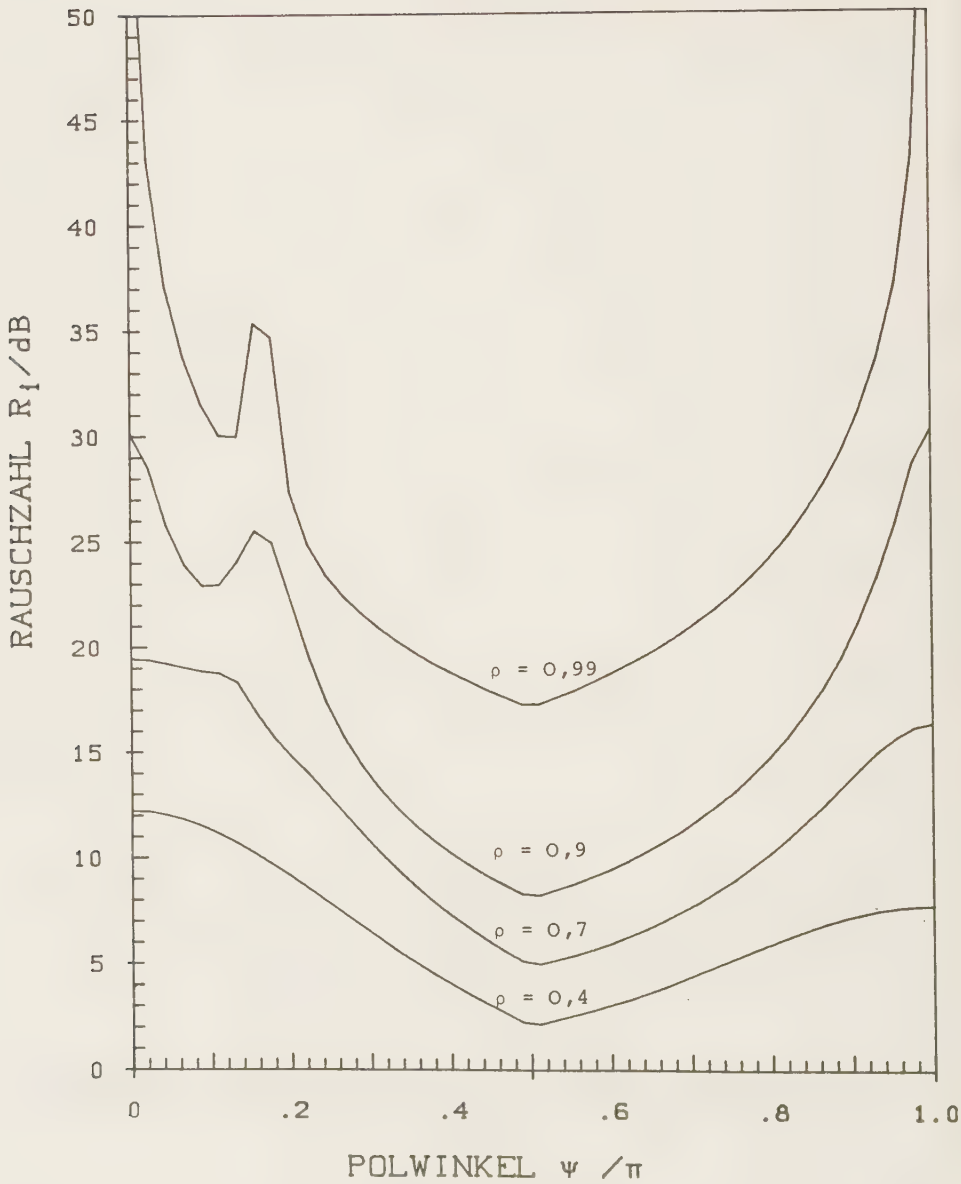


Bild 5.7: Pol-Nullstellen-Test der zweiten Struktur nach Meerkötter und Wegener

Nullstellen: $z_{01,2} = e^{\pm j\pi/6}$

Polstellen: $z_{\infty 1,2} = \rho e^{\pm j\psi}$

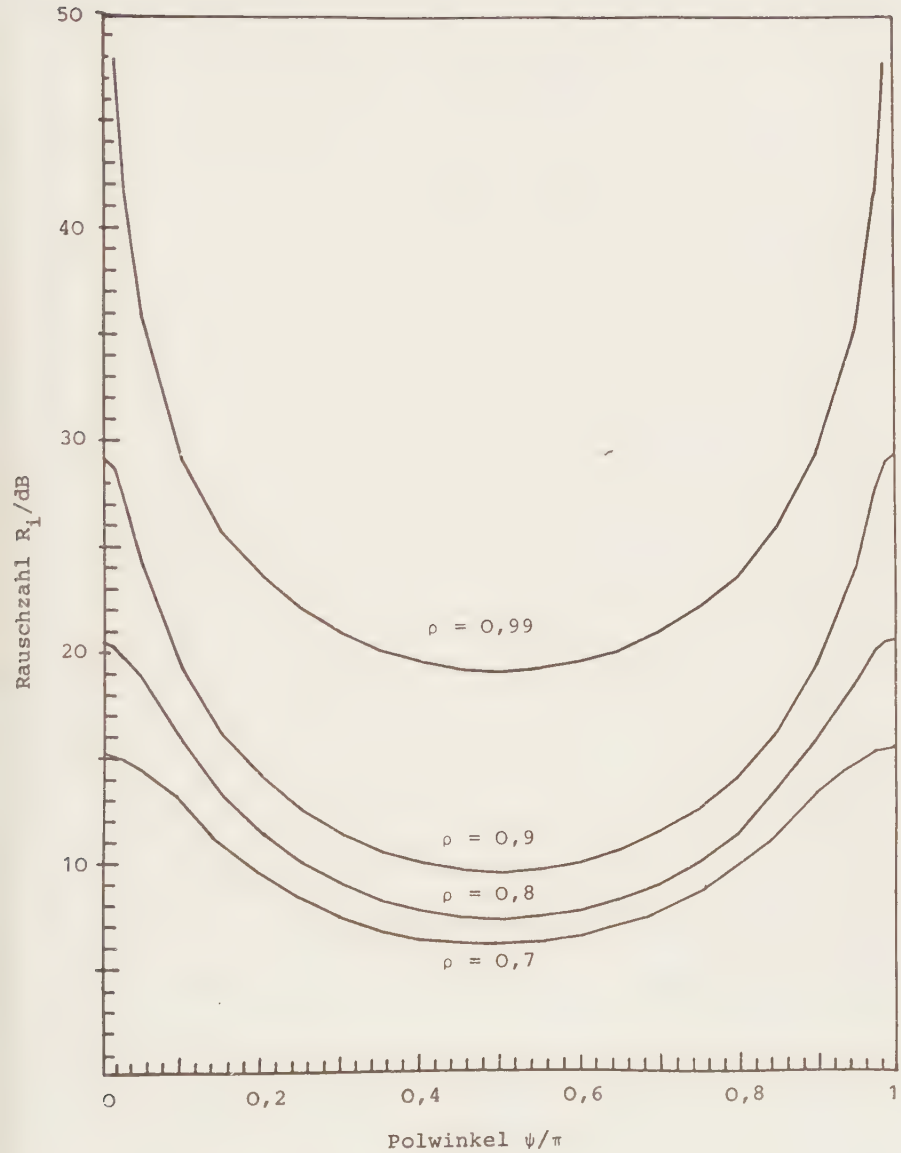


Bild 5.8: Pol-Nullstellentest eines Blockes zweiten Grades in 1. kanonischer Form

Nullstellen: $z_{0,1,2} = e^{\pm j\pi/6}$

Polstellen: $z_{\infty 1,2} = \rho e^{\pm j\psi}$

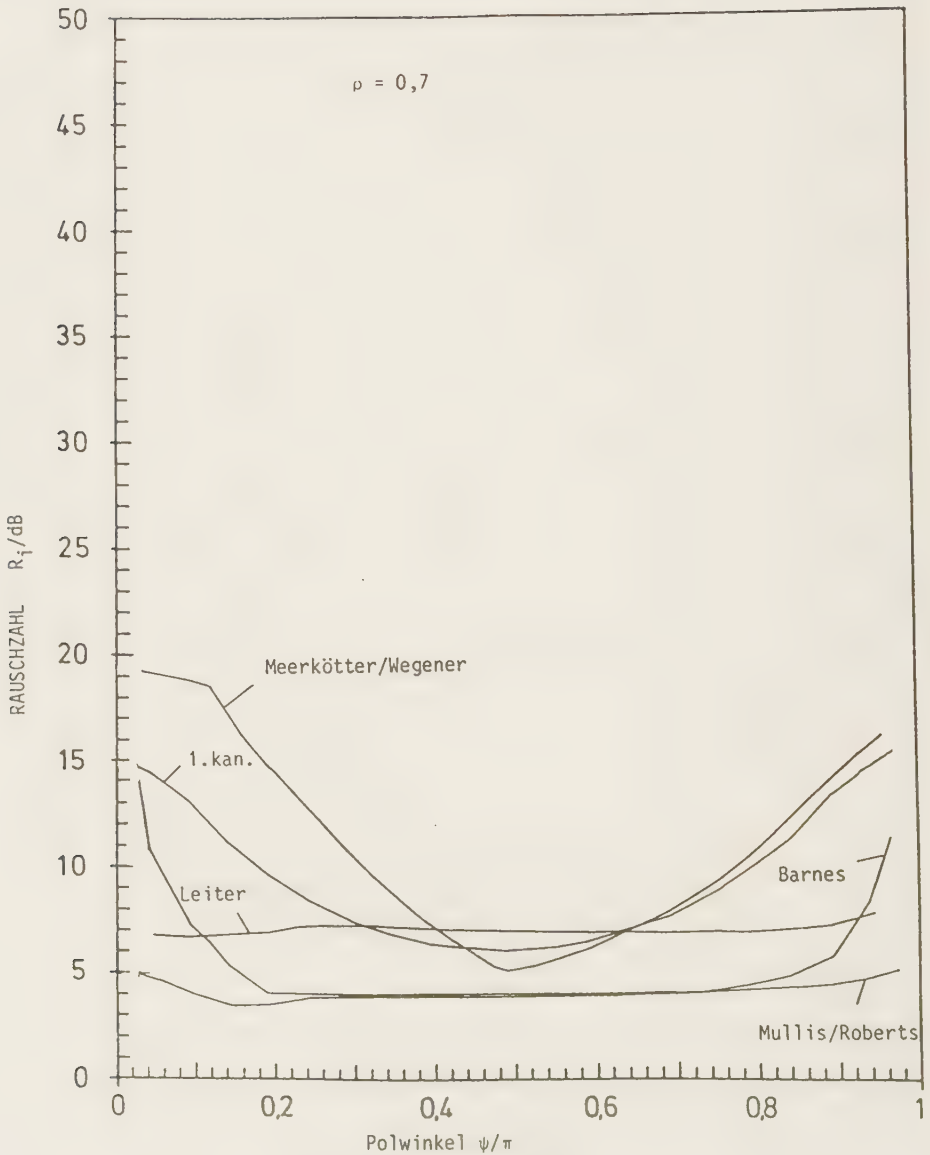


Bild 5.9: Pol-Nullstellen-Test

Vergleich der Strukturen

Nullstellen: $z_{01,2} = e^{\pm j\pi/6}$

Polstellen: $z_{\infty 1,2} = \rho e^{\pm j\psi}$; $\rho = 0,7$

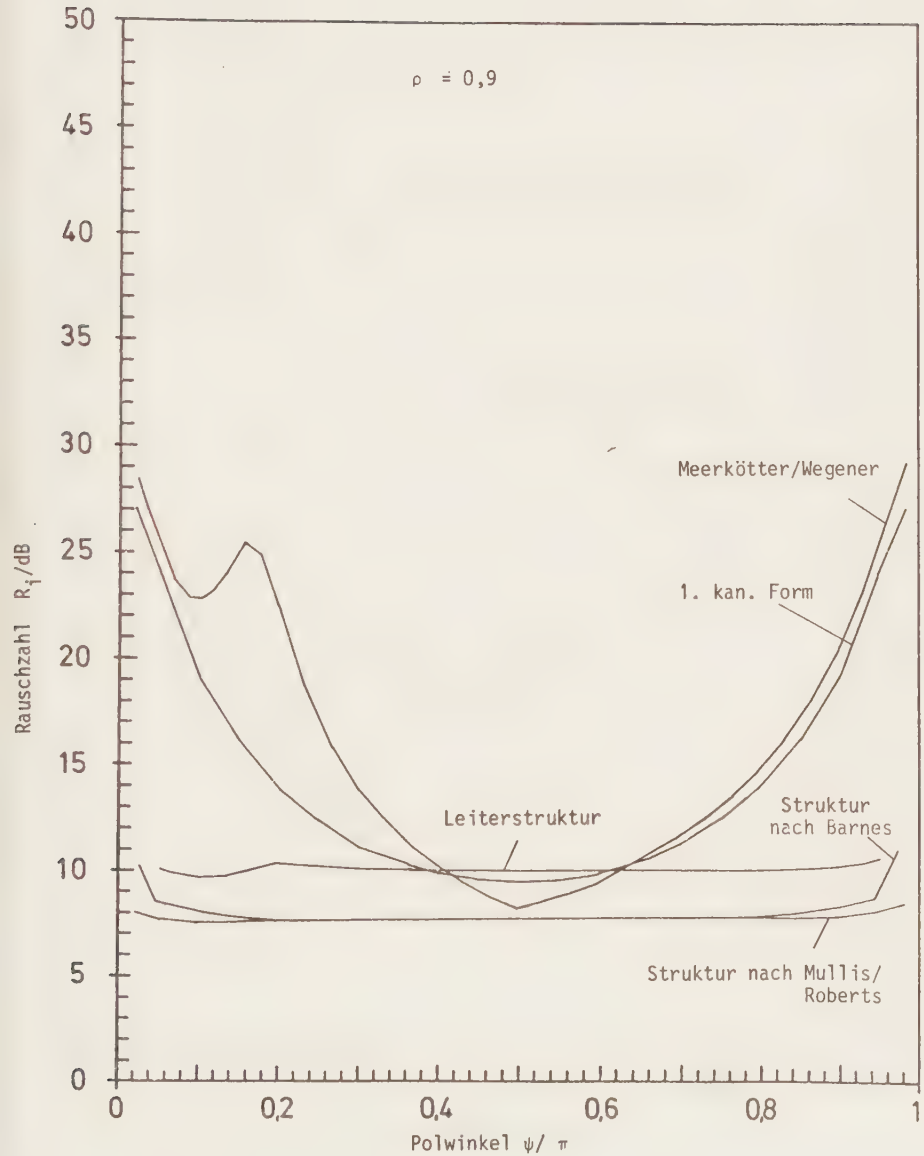


Bild 5.10: Pol-Nullstellen-Test für Filter 2-ten Grades
Vergleich der Strukturen

Nullstellen: $z_{01,2} = e^{\pm j\pi/6}$

Polstellen: $z_{\infty 1,2} = \rho e^{\pm j\psi}$; $\rho = 0.9$

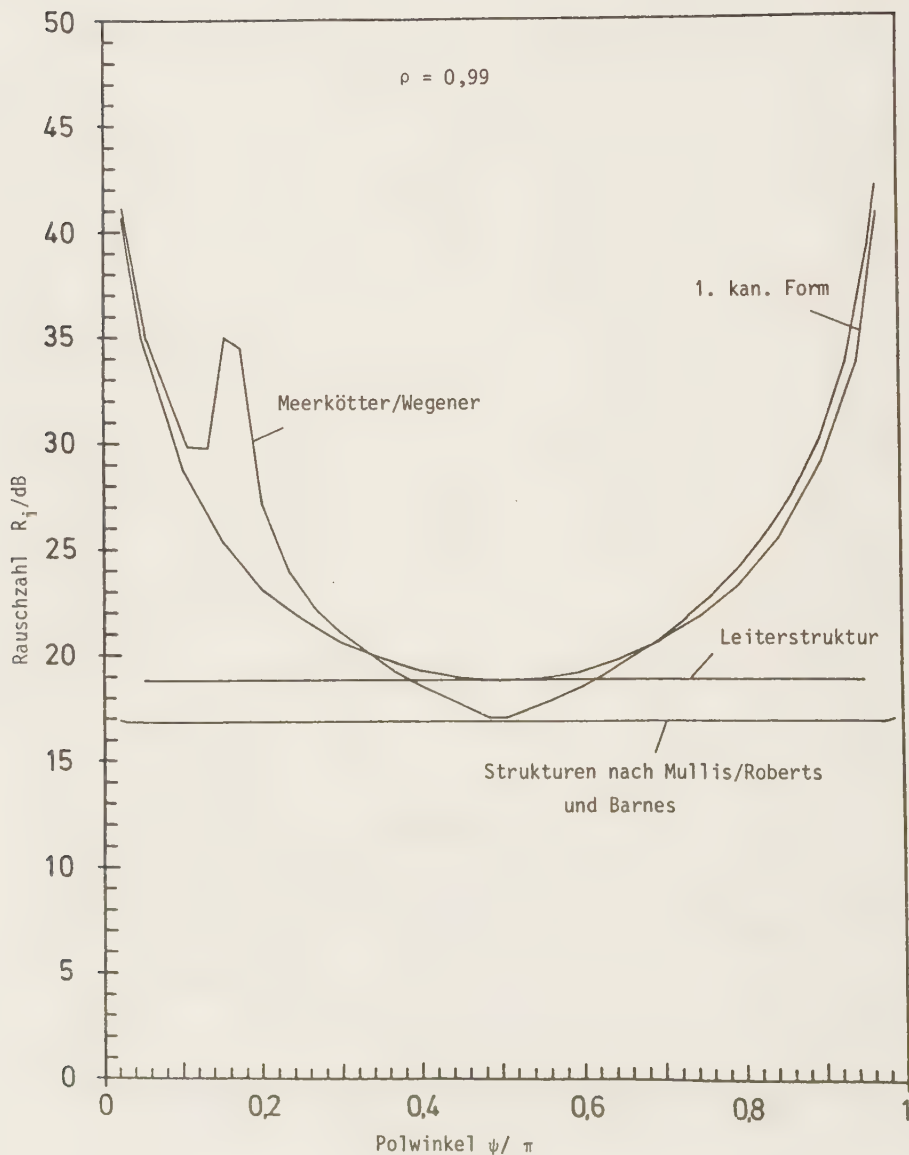


Bild 5.11: Pol-Nullstellen-Test für Filter 2-ten Grades, Vergleich der Strukturen

Nullstellen: $z_{01,2} = e^{\pm j\pi/6}$

Polstellen: $z_{\infty 1,2} = \rho e^{\pm j\psi}$; $\rho = 0,99$

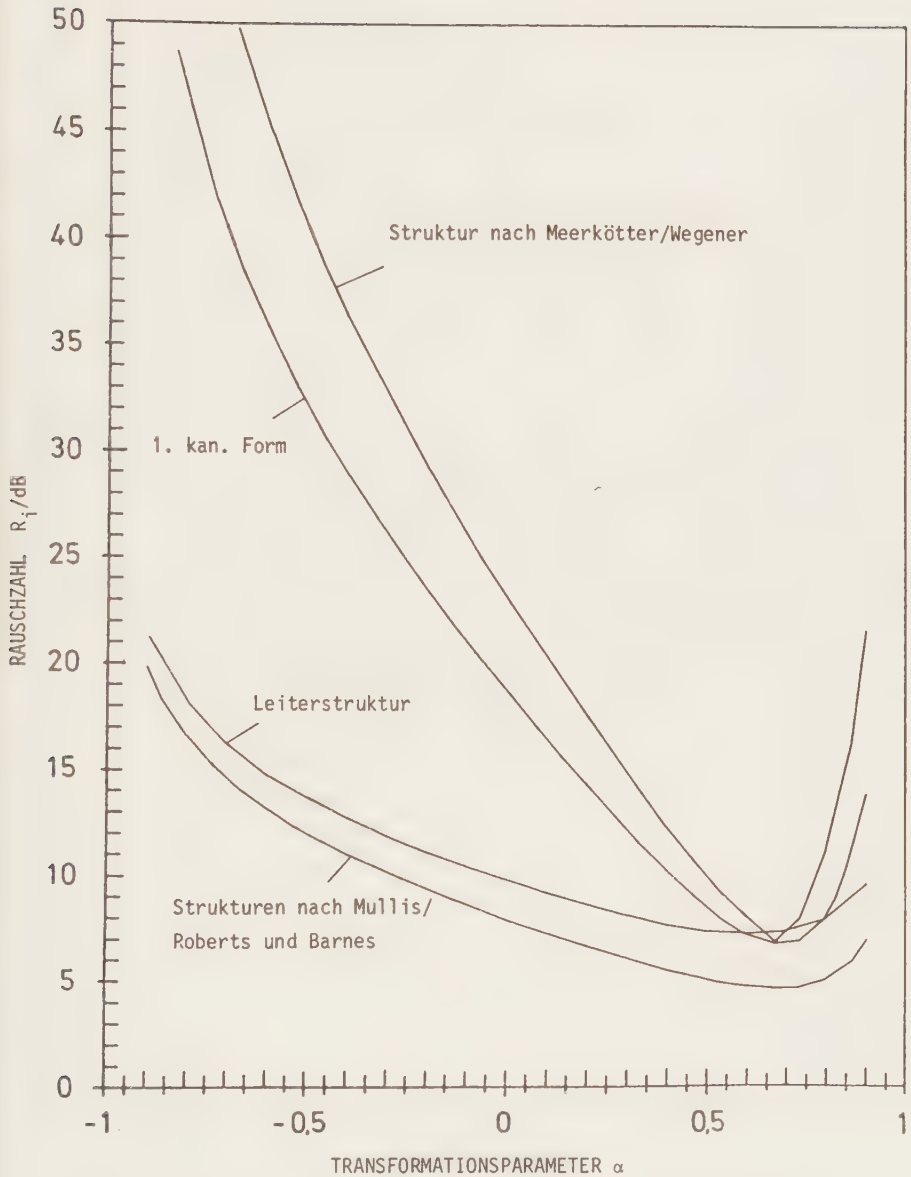


Bild 5.12: Rundungsgeräusch von Filtern 2-ten Grades bei Allpaßtransformation

Prototyp mit $z_{01,2} = e^{\pm j\pi/6}$, $z_{\infty 1,2} = 0,9 e^{\pm j\pi/9}$

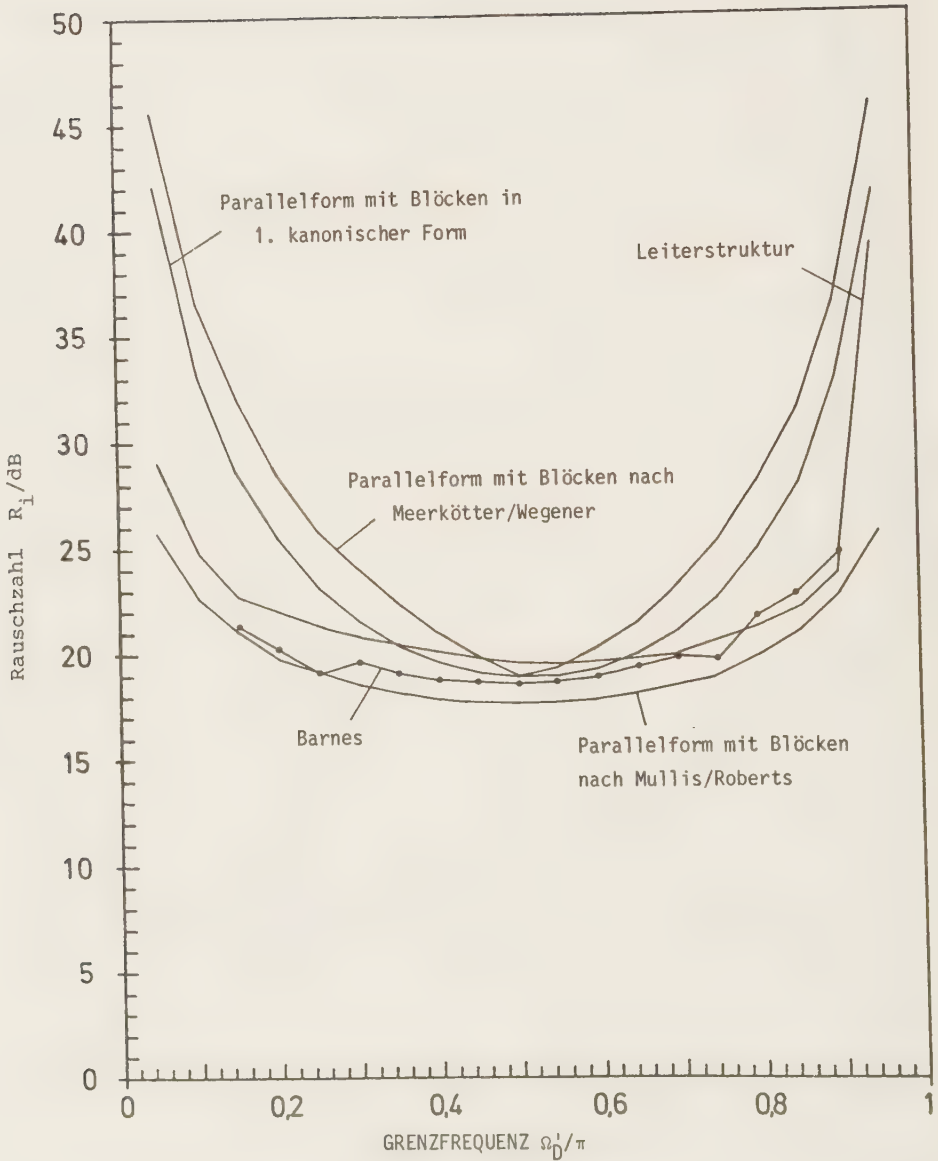


Bild 5.13: Rundungsgeräusch eines Tiefpasses 10-ten Grades bei Variation der Grenzfrequenz ω_D' durch Allpaßtransformation

5.4 Granulare Grenzzyklen

5.4.1 Meßergebnisse

Untersucht wurden Filterblöcke zweiten Grades mit unterschiedlichen Pol-Lagen. Die Pole und Nullstellen sind

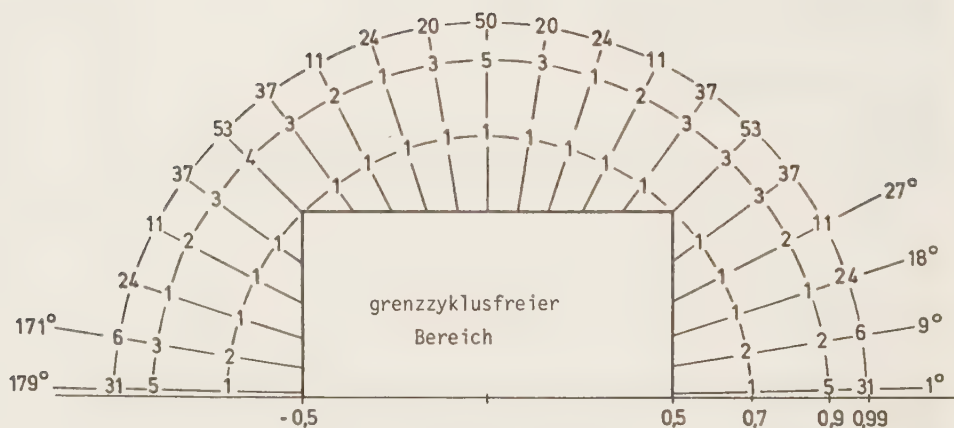
$$z_{\infty 1,2} = \rho e^{\pm j\psi}, \rho \text{ und } \psi \text{ werden variiert}$$

$$z_{01,2} = e^{\pm j\pi/6}$$

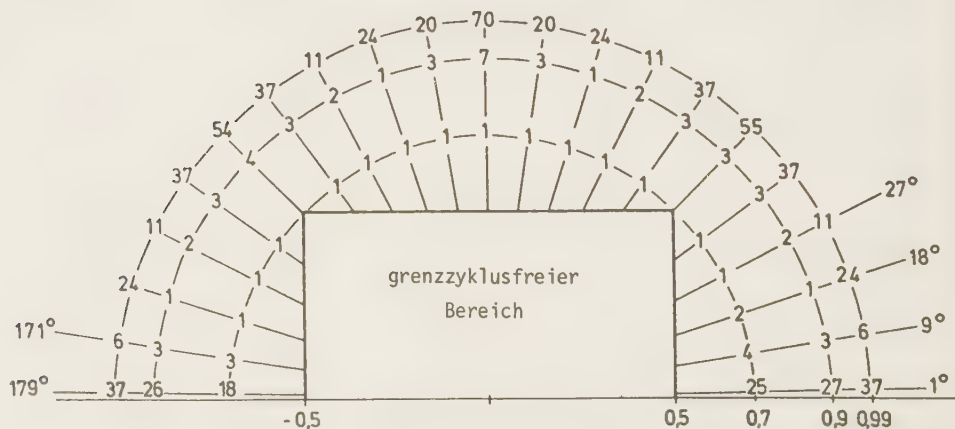
Die Ergebnisse wurden teils durch Simulation am Allzweckrechner PDP 11/45 (Strukturen nach Mullis/Roberts und Barnes) und im übrigen durch Messung am Mikrosignalprozessor MSP 82 ermittelt.

Eine Zusammenstellung der Resultate findet sich in den Bildern 5.14 bis 5.21^{*)}. Die Vergleichswerte für einen Block zweiten Grades in erster kanonischer Form wurden aus /51/ übernommen (Bilder 5.18 bis 5.21).

^{*)} In den Bildern 5.14 bis 5.17 wurde eine Darstellung in der z-Ebene gewählt, wobei an der jeweiligen Polstellenlage die maximale Grenzzyklus-Amplitude bezogen auf das Quantisierungsintervall Q angegeben ist.

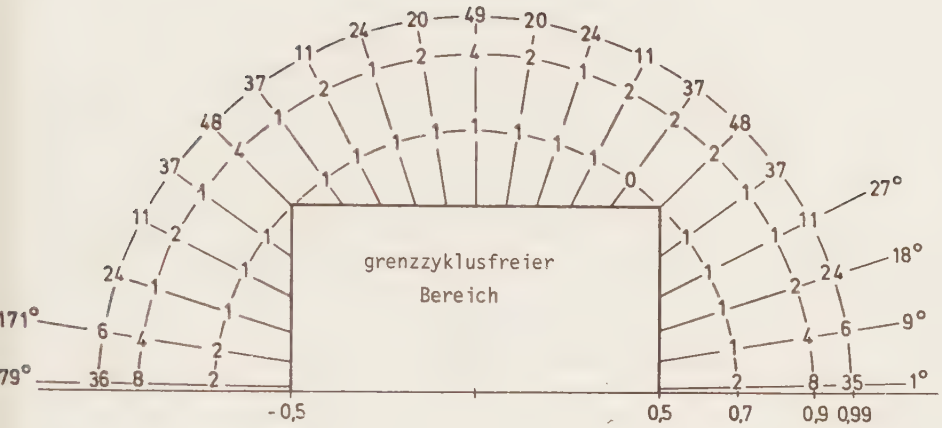


a) größte Amplitude der Zustandsvariablen

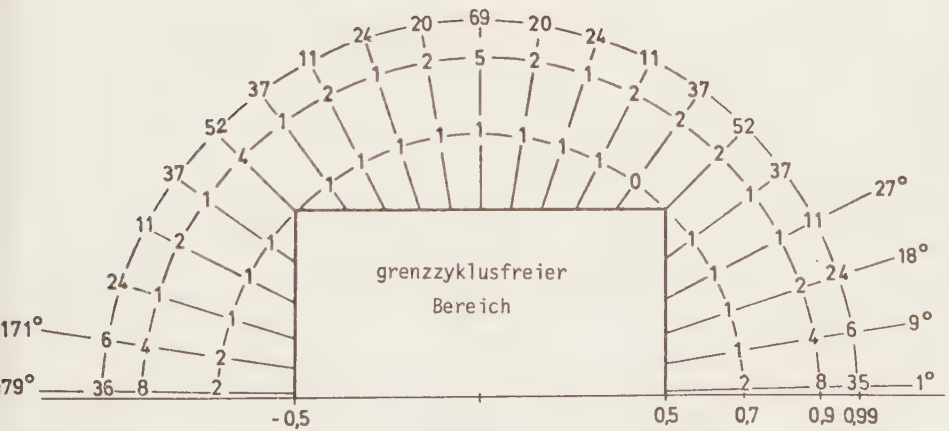


b) größte Amplitude am Filterausgang

Bild 5.14: Grenzzyklus-Amplituden von Filtern 2-ten Grades nach Barnes (unterschiedliche Pollagen, Nullstellen $z_{01,2} = e^{\pm j\pi/6}$).



a) größte Amplitude der Zustandsvariablen



b) größte Amplitude am Filterausgang

Bild 5.15: Grenzzyklus-Amplituden von Filtern zweiten Grades nach Mullis und Roberts (unterschiedliche Pollagen, Nullstelle $z_{01,2} = e^{\pm j\pi/6}$)

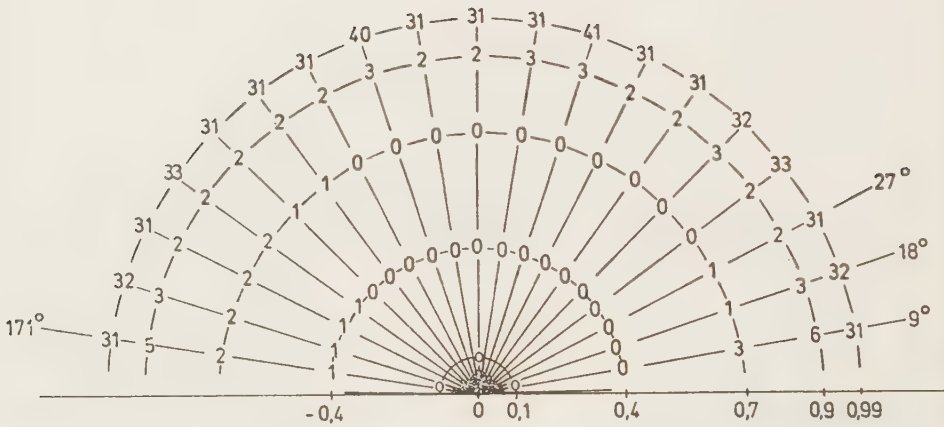


Bild 5.16: Grenzzyklus-Amplituden (Filterausgang) der Leiterstruktur nach Markel und Gray (4-Multiplizierer-Form, Grad 2 mit Nullstellen $z_{0,1,2} = e^{\pm j\pi/6}$)

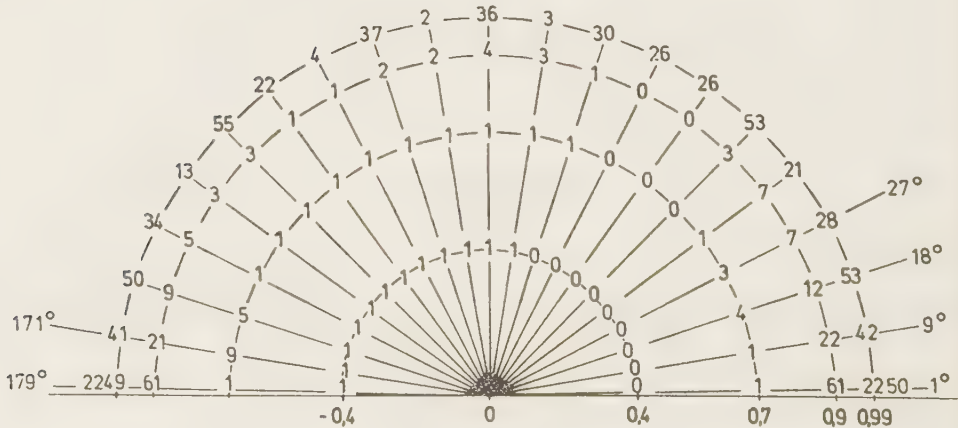


Bild 5.17: Grenzzyklus-Amplituden (Filterausgang) von Filtern zweiten Grades nach Meerkötter/Wegener (unterschiedliche Pol-Lagen, Nullstelle $z_{0,1,2} = e^{\pm j\pi/6}$)

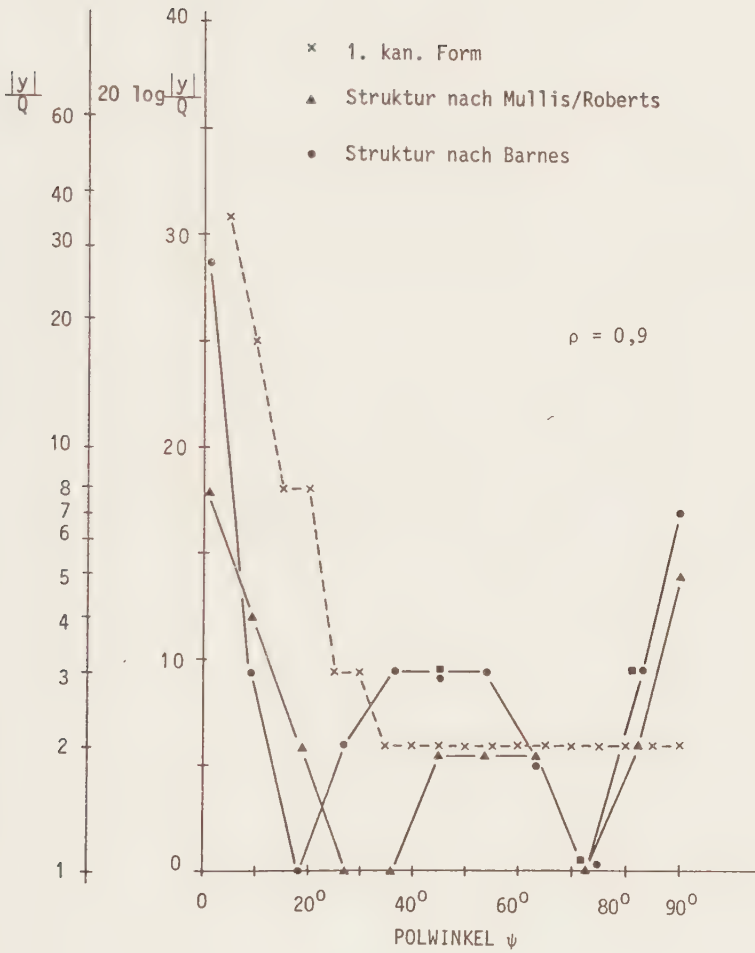


Bild 5.18: Grenzzyklus-Amplituden von Filtern 2-ten Grades (Filterausgang)

Polstellen: $z_{\infty 1,2} = 0,9 e^{\pm j\psi}$

Nullstellen: $z_{0 1,2} = e^{\pm j30^\circ}$

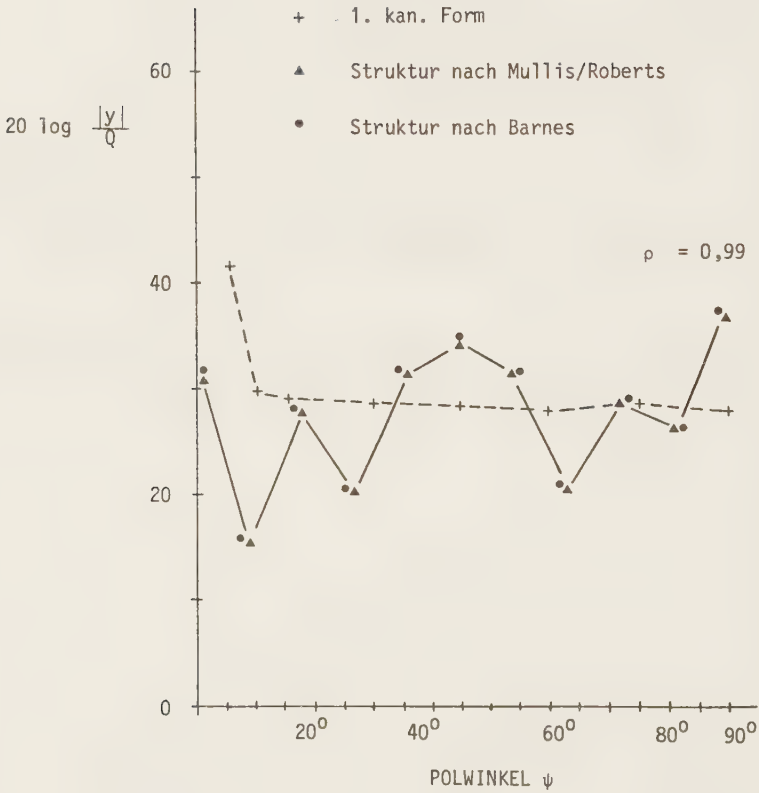


Bild 5.19: Grenzzyklus-Amplituden von Filtern 2-ten Grades (Filterausgang)

Polstellen: $z_{\infty 1,2} = 0,99 e^{\pm j\psi}$

Nullstellen: $z_{0 1,2} = e^{\pm j30^\circ}$

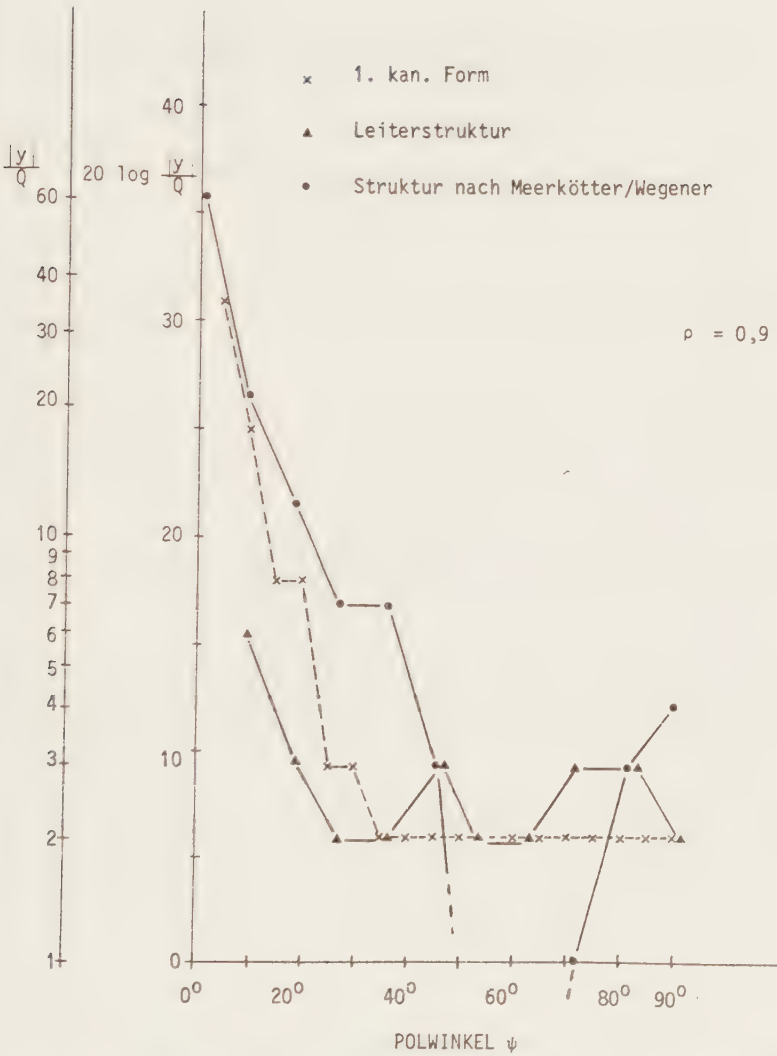


Bild 5.20: Grenzyklus-Amplituden von Filtern 2-ten Grades (Filterausgang)

Polstellen: $z_{\infty 1,2} = 0,9 e^{\pm j\psi}$

Nullstellen: $z_{0 1,2} = e^{\pm j30^\circ}$

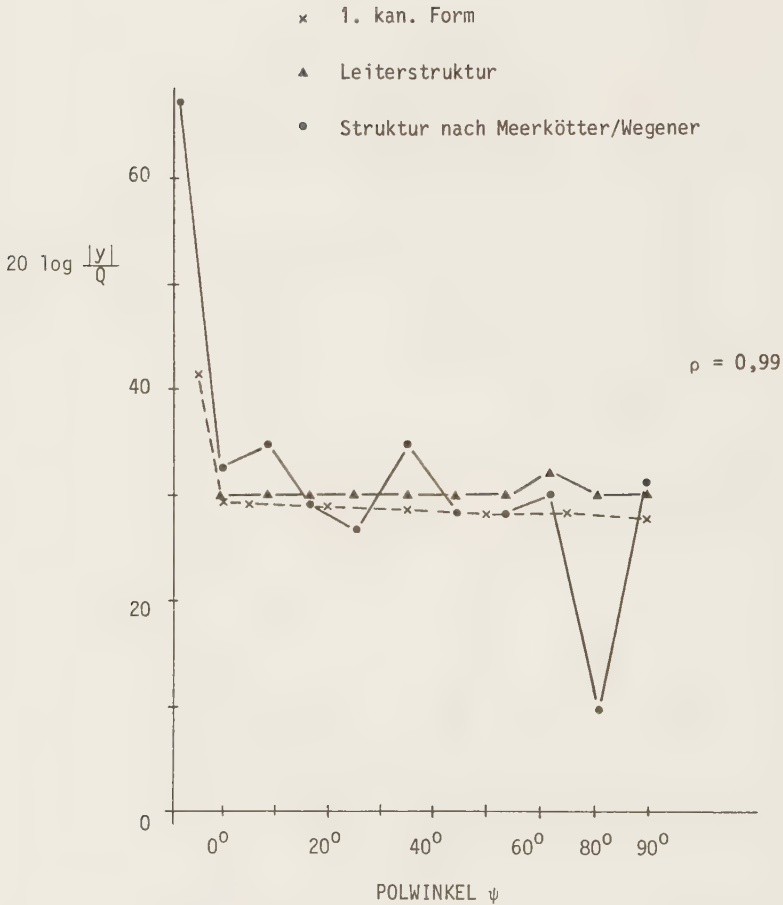


Bild 5.21: Grenzzzyklus-Amplituden von Filtern 2-ten Grades (Filterausgang)

Polstellen: $z_{\infty 1,2} = 0,99 e^{\pm j\psi}$

Nullstellen: $z_{0 1,2} = e^{\pm j30^\circ}$

5.4.2 Theoretische Abschätzung der Grenzzyklus-Amplituden

Die folgenden Überlegungen beziehen sich auf Strukturen mit der Eigenschaft

$$\|\underline{A} \underline{x}\|_{\underline{G}} \leq \xi \|\underline{x}\|_{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n, \quad 0 \leq \xi < 1 \quad (5-3)$$

Ferner wird, wie üblich, gefordert, daß die Rundung am Eingang der Zustandspeicher (nach Zusammenfassen der Teilprodukte) erfolgen soll.

Die genannten Bedingungen werden z.B. von den Filtern nach Mullis/Roberts und Barnes erfüllt, wobei in (5-3) $\xi = \max_v \{|\lambda_v|\}$ gilt und λ_v die Eigenwerte von $\underline{A}$ bezeichnet.

Das Ziel ist nun, den Bereich im Zustandsraum abzugrenzen, in welchem sich der Zustandsvektor bei granularen Grenzzyklen befindet.

Zunächst wird ein allgemeines Filter n-ten Grades (n beliebig) betrachtet. Das Ergebnis läßt sich dann auf den Fall $n=2$ spezialisieren, woraus sich u.a. ein Gebiet grenzzyklusfreier Polstellen-Lagen ableitet.

Bei verschwindender Erregung $v(k) \equiv 0$ gilt unter Berücksichtigung der Rundungsfehler

$$[\underline{A} \underline{x}(k)]_Q = \underline{x}(k+1) = \underline{A} \underline{x}(k) + \underline{\Delta}(k), \quad (5-4)$$

wobei

$$\underline{\Delta}(k) = [\Delta_1(k), \Delta_2(k), \dots, \Delta_n(k)]^T \quad (5-5)$$

mit $|\Delta_v(k)| \leq \frac{Q}{2}$ (Q = Quantisierungsintervall)

den Fehlervektor bezeichnet. Damit folgt

$$\begin{aligned}
 ||\underline{x}(k+1)||_{\underline{G}}^2 &= [\underline{A} \underline{x}(k) + \underline{\Delta}(k)]^T \underline{G} [\underline{A} \underline{x}(k) + \underline{\Delta}(k)] \\
 &= \underbrace{[\underline{A} \underline{x}(k)]^T \underline{G} [\underline{A} \underline{x}(k)]}_{= ||\underline{A} \underline{x}(k)||_{\underline{G}}^2} + 2 \underbrace{[\underline{A} \underline{x}(k)]^T \underline{G} \underline{\Delta}(k)}_{:= \tilde{\underline{x}}(k+1)} + \underline{\Delta}^T(k) \underline{G} \underline{\Delta}(k) .
 \end{aligned}$$

Mit

$$\tilde{\underline{x}} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix}, \quad \underline{G} = \text{Diag} \{G_1, G_2, \dots, G_n\},$$

$$|\Delta_v(k)| \leq \frac{Q}{2}$$

und bei Beachtung von (5-3) gewinnt man

$$\begin{aligned}
 ||\underline{x}(k+1)||_{\underline{G}}^2 &\leq \xi^2 ||\underline{x}(k)||_{\underline{G}}^2 + 2 \frac{Q}{2} \max_{\underline{x}_\mu} \{ \sum_{\mu=1}^n G_\mu \tilde{x}_\mu(k+1) \} \\
 &\quad + \frac{Q^2}{4} \sum_{\mu=1}^n G_\mu
 \end{aligned} \tag{5-6}$$

Wie in Anhang A7 gezeigt wird, gilt

$$\max_{\underline{x}_\mu} \{ \sum_{\mu=1}^n G_\mu \tilde{x}_\mu(k+1) \} \leq \xi ||\underline{x}(k)||_{\underline{G}} \sqrt{\sum_{\mu=1}^n G_\mu} . \tag{5-7}$$

Damit geht (5-6) über in

$$||\underline{x}(k+1)||_{\underline{G}}^2 \leq \xi^2 ||\underline{x}(k)||_{\underline{G}}^2 + 2 \frac{Q}{2} \xi ||\underline{x}(k)||_{\underline{G}} \sqrt{\sum_{\mu=1}^n G_\mu} + \frac{Q^2}{4} \sum_{\mu=1}^n G_\mu$$

bzw.

$$||\underline{x}(k+1)||_{\underline{G}}^2 \leq \left[\xi ||\underline{x}(k)||_{\underline{G}} + \frac{Q}{2} \sqrt{\sum_{\mu=1}^n G_\mu} \right]^2$$

bzw.

$$||\underline{x}(k+1)||_{\underline{G}} \leq \xi ||\underline{x}(k)||_{\underline{G}} + \frac{Q}{2} \sqrt{\sum_{\mu=1}^n G_\mu} \tag{5-8}$$

Aus (5-8) und der Forderung

$$||\underline{x}(k+1)||_{\underline{G}} < ||\underline{x}(k)||_{\underline{G}} \quad (5-9a)$$

gewinnt man

$$||\underline{x}(k)||_{\underline{G}} > Q \frac{\frac{1}{2} \sqrt{\sum_{\mu=1}^n G_{\mu}}}{1 - \xi} \quad (5-9b)$$

Das bedeutet:

Für ZV-Vektoren $\underline{x}(k)$, welche die Bedingung (5-9b) erfüllen, erfolgt immer eine Abnahme der Norm im nächsten Filterschritt. Umgekehrt ist eine Zunahme oder Erhaltung der Norm nur für Vektoren $\underline{x}(k)$ mit

$$||\underline{x}(k)||_{\underline{G}} \leq Q \frac{\frac{1}{2} \sqrt{\sum_{\mu=1}^n G_{\mu}}}{1 - \xi} \quad (5-10)$$

möglich.

Wir fragen nun nach der größtmöglichen ZV-Norm im Grenzyklusfall (Abschätzung einer oberen Schranke). Der zu dieser größten Norm gehörende ZV-Vektor sei $\underline{x}(k=k_0)$. Dann muß der im Grenzyklus vorhergehende ZV-Vektor $\underline{x}(k_0-1)$ die Bedingung (5-10) erfüllen, da andernfalls

$$||\underline{x}(k_0)||_{\underline{G}} < ||\underline{x}(k_0-1)||_{\underline{G}}$$

gelten würde.

Mit (5-8) folgt die Abschätzung

$$||\underline{x}(k_0)||_{\underline{G}} \leq \frac{Q}{2} \sqrt{\sum_{\mu=1}^n G_{\mu}} \left[\frac{\xi}{1-\xi} + 1 \right] = \frac{Q}{2} \frac{\sqrt{\sum_{\mu=1}^n G_{\mu}}}{1-\xi} \quad (5-11)$$

Wie ein Vergleich zeigt, ist also bereits mit (5-10) eine obere Schranke für die ZV-Vektor-Norm von Grenzyklen gegeben.

Bei den hier interessierenden Filter-Strukturen (Mullis/Roberts, Barnes) war $\xi = \max_v \{|\lambda_v|\}$, so daß sich (5-10) zunächst als

$$||\underline{x}(k)||_{\underline{G}} \leq Q \frac{\frac{1}{2} \sqrt{\sum_{\mu=1}^n G_{\mu}}}{1 - \max\{|\lambda_v|\}} \quad (5-12)$$

schreiben läßt. Eine weitere Spezialisierung auf Blöcke zweiten Grades mit konjugiert komplexen Polstellen $\lambda_{1,2} = \rho e^{\pm j\psi}$, $|\lambda_{1,2}| = \rho$, liefert

$$||\underline{x}(k)||_{\underline{G}} \leq Q \frac{\frac{1}{2} \sqrt{G_1 + G_2}}{1 - \rho} \quad (5-13)$$

Setzt man schließlich noch normale A-Matrizen voraus, so erhält man wegen $\underline{G} = \underline{E}$ bzw. $G_1 = G_2 = 1$:

$$||\underline{x}(k)||_{\underline{E}} \leq Q \frac{1}{\sqrt{2} (1-\rho)} \quad (5-14a)$$

Für die Komponenten des ZV-Vektors gilt hier als weitere Abschätzung nach oben

$$x_{1,2}(k) \leq Q \frac{1}{\sqrt{2} (1-\rho)} \quad (5-14b)$$

(5-14b) stellt natürlich eine schwächere Aussage als (5-14a) dar. Nach (5-14a) gilt ja

$$\sqrt{x_1^2(k) + x_2^2(k)} \leq Q \frac{1}{\sqrt{2} (1-\rho)} .$$

Nehmen wir nun an, daß der ZV-Vektor $\underline{x}(k_0)$ mit der größten Norm betragsmäßig gleiche Komponenten aufweist, so folgt

$$\sqrt{2x_1^2(k_0)} = \sqrt{2x_2^2(k_0)} \leq Q \frac{1}{\sqrt{2} (1-\rho)}$$

bzw.

$$|x_1(k_0)| = |x_2(k_0)| \leq Q \frac{1}{2(1-\rho)} . \quad (5-15)$$

D.h. in diesem Spezialfall wurde mit (5-14b) die obere Grenze um den Faktor $\sqrt{2}$ überschätzt. (5-15) deckt sich im übrigen mit dem von Jackson /26/ mit Hilfe seines "Effektive-Value"-Modells ermittelten Ergebnis, wenn die Polstellen in der Nähe bzw. auf der imaginären oder reellen Achse liegen. Jackson nimmt zwar Rundung

nach jeder Multiplikation an, doch entspricht dies einer Rundung nach Zusammenfassung der Teilprodukte, wenn die Polstellen (= Eigenwerte von A) rein reell bzw. imaginär sind.

Obwohl die Beziehungen (5-14a,b) für normale A-Matrizen hergeleitet wurden, erweisen sie sich jedoch auch im Fall der nachträglich L_∞ -skalierten Strukturen nach Mullis und Roberts sowie nach Barnes als brauchbar. Dazu werten wir (5-14b) für einige Polradien ρ aus und vergleichen mit den Simulationsergebnissen der Bilder 5.14, 5.15, wobei die für den jeweiligen Polradius größte Grenzzykus-Amplitude herausgegriffen wird (Tabelle 5.2).

ρ	$Q \frac{1}{\sqrt{2} (1-\rho)}$	$\max\{ x_{1,2} \}$ (Simulation)	
		Mullis/Roberts	Barnes
0,7	2,36 Q	2 Q	2 Q
0,9	7,07 Q	8 Q	5 Q
0,99	70,71 Q	49 Q	53 Q

Tabelle 5.2: Vergleich von Simulationsergebnissen und theoretischer Abschätzung

Im folgenden soll gezeigt werden, daß Systeme 2-ten Grades mit

$$\|\underline{A} \underline{x}\|_{\underline{G}} \leq \max_v \{|\lambda_{1,2}|\} \|\underline{x}\|_{\underline{G}}$$

und $|\lambda_{1,2}| < \frac{1}{2}$

grenzzykusfrei sind.

Wir gehen zunächst von der Abschätzung (5-12) aus und erhalten für $n = 2$

$$\|\underline{x}(k)\|_{\underline{G}} \leq Q \frac{\frac{1}{2} \sqrt{G_1 + G_2}}{1 - \max\{|\lambda_{1,2}|\}} \quad (5-16)$$

Mit $|\lambda_{1,2}| < \frac{1}{2}$ folgt

$$||\underline{x}(k)||_{\underline{G}} < Q \sqrt{G_1 + G_2} \quad (5-17a)$$

bzw.

$$G_1 x_1^2(k) + G_2 x_2^2(k) < Q^2 [G_1 + G_2] \quad (5-17b)$$

Der ZV-Vektor $\underline{x}(k)$ beschreibt demnach einen Punkt im Innern einer Ellipse mit den Halbachsen

$$Q \sqrt{\frac{G_1 + G_2}{G_1}}$$

in x_1 -Richtung und

$$Q \sqrt{\frac{G_1 + G_2}{G_2}}$$

in x_2 -Richtung. O.E.d.A. wird nun $G_1 \geq G_2$ angenommen. Dann liegt $\underline{x}(k)$ im Innern des Kreises in Bild 5.22. (Die x_1 - und x_2 -Achse sei hierbei so skaliert, daß die Ellipse gemäß (5-17) in einen Kreis übergeht. Im folgenden wird gezeigt, daß die Norm des ZV-Vektors in jedem Filterschritt trotz Rundung abnimmt.

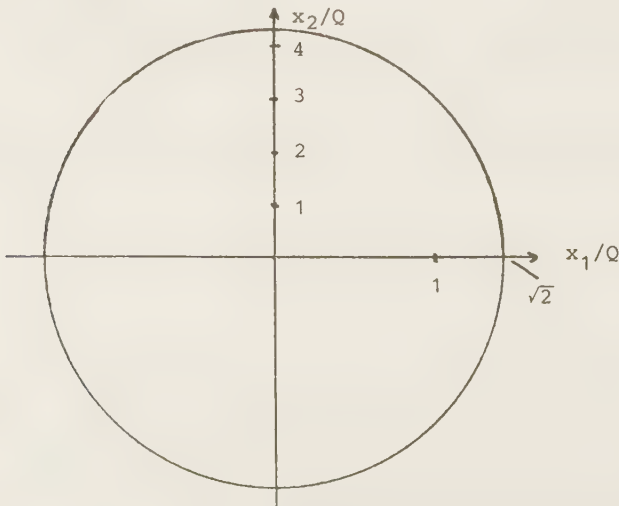


Bild 5.22:

Dazu nehmen wir an, daß $\underline{x}(k)$ auf dem in Bild 5.23 gezeichneten Kreis I (Punkte konstanter Norm) liegt.

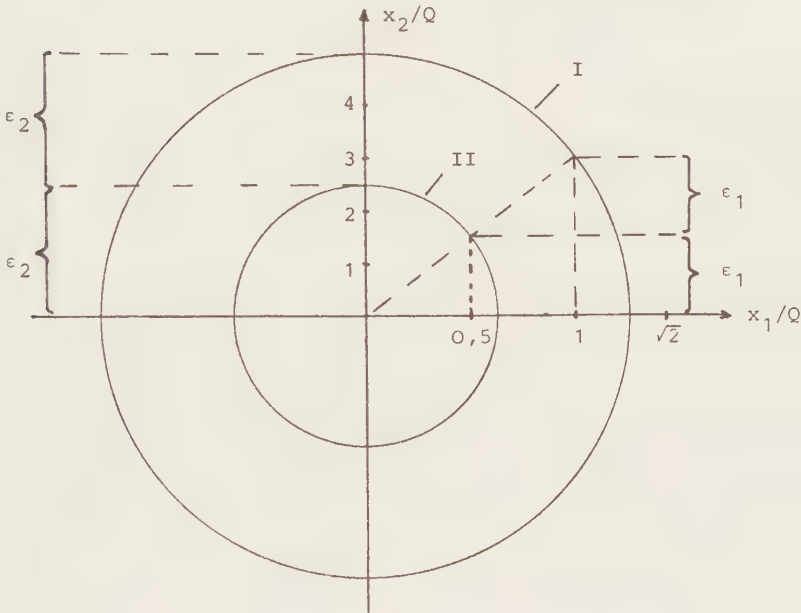


Bild 5.23:

Wegen $||\underline{A} \underline{x}(k)||_{\underline{G}} = ||\tilde{\underline{x}}(k+1)||_{\underline{G}} < \frac{1}{2} ||\underline{x}(k)||_{\underline{G}}$ liegt der ideal berechnete neue ZV-Vektor im Innern von Kreis II.

Fall 1:

Ist die Komponente $|\tilde{x}_1(k+1)| \geq \frac{Q}{2}$, so wird aufgerundet und es gilt $|x_1(k+1)| = Q$. Dies ist aber nur möglich, wenn gleichzeitig $|\tilde{x}_2(k+1)| < \epsilon_1 \cdot Q$ ist. Nach Rundung erhält man

$$|x_2(k+1)| < 2\epsilon_1 Q .$$

Der neue ZV-Vektor liegt damit innerhalb Kreis I, was einer Abnahme der Norm entspricht.

Fall 2: $|\tilde{x}_1(k+1)| < \frac{Q}{2}$

Die Komponente $\tilde{x}_1$ wird abgerundet, so daß $x_1(k+1) = 0$ ist.
 $\tilde{x}(k+1)$ liegt innerhalb Kreis II. Damit folgt

$$|\tilde{x}_2(k+1)| < \varepsilon_2 Q$$

und nach Rundung

$$|x_2(k+1)| < 2\varepsilon_2 Q.$$

Der Vektor $\underline{x}(k+1)$ liegt innerhalb Kreis I.

Damit ist der Beweis für obige Behauptung erbracht.

Der Bereich grenzzyklusfreier Polstellenlagen wird durch den Kreis in Bild 5.24 begrenzt, wobei auch reelle Pole erfaßt werden. Das Ergebnis gilt für alle Systeme zweiten Grades, deren $\underline{A}$ -Matrix durch Ähnlichkeitstransformation mit einer reellen Diagonalmatrix $\underline{G}^{1/2}$ normal gemacht werden können, wobei keine Voraussetzungen bzgl. $\underline{G}^{1/2}$ getroffen wurden.

Für den Fall normaler $\underline{A}$ -Matrizen mit konjugiert komplexen Eigenwerten geben Barnes und Shinnaka /15/ einen erweiterten Bereich an (Quadrat in Bild 5.24). Der theoretische Beweis wird aber in /15/ nicht geführt. Es wird lediglich gesagt, daß ein sehr langer und wenig eleganter Beweis existiert.

Es ist zu vermuten (Messungen haben dies bestätigt), daß das Ergebnis von Barnes auch für den oben behandelten allgemeineren Fall gilt.

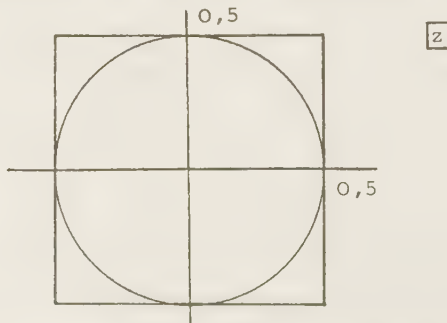


Bild 5.24: Bereich grenzzyklusfreier Polstellen-Lagen

6. FILTER - REALISIERUNG DURCH BLOCKVERARBEITUNG

6.1 Vorbemerkungen

Zeman und Lindgren schlugen in /40/ und /41/ eine Realisierung rekursiver Filter durch segmentweise Verarbeitung der Eingangsfolge vor. Der Algorithmus bringt eine Verminderung von Rundungsgeräusch und Grenzyklus-Amplituden mit sich und eignet sich darüber hinaus gut für eine Abtastraten-Reduktion am Filterausgang. Das Verfahren basiert auf der Zustandsraum-Darstellung, wobei prinzipiell beliebige Filterstrukturen als Grundlage herangezogen werden können. In vielen Fällen verringert sich der arithmetische Aufwand gegenüber der Grundstruktur. Dies gilt insbesondere auch für die in Kap. 3 behandelten überlaufstabilen und rauscharmen Filter nach Mullis/Roberts und Barnes (Normalform-Struktur), deren Stabilitätseigenschaften im übrigen erhalten bleiben.

Barnes schlug zusammen mit Shinnaka und Leung im Prinzip das gleiche Verfahren vor, wobei speziell die Normalform-Struktur als Basis herangezogen wurde. Ihre Ergebnisse finden sich in /15/, /16/, /42/ und /43/.

Im folgenden soll der genannte Filteralgorithmus und seine Eigenschaften bzgl. des Rundungsrauschens zunächst kurz erläutert werden. Die Ausführungen orientieren sich im wesentlichen an /40/ und /42/. Ergänzend werden allgemeine Betrachtungen bzgl. Überlaufstabilität und granularer Grenzyklen angestellt. Dies entspricht einer Erweiterung der von Barnes, Leung und Shinnaka in /15/ und /42/ für Basisstrukturen in Normalform gefundenen Aussagen. Schließlich werden Meßergebnisse vorgestellt, welche bei der praktischen Realisierung mit Hilfe des am Institut für Nachrichtentechnik entwickelten Signalprozessors MSP 82 /44/ gewonnen wurden.

6.2_Der_Filteralgorithmus

Die Zustandsraum-Beschreibung einer beliebigen Filterstruktur mit einem Eingang und einem Ausgang sei gegeben als

$$\underline{x}(k+1) = \underline{A} \underline{x}(k) + \underline{b} v(k) \quad (6-1a)$$

$$y(k) = \underline{c}^T \underline{x}(k) + d v(k) \quad (6-1b)$$

$$\underline{A} = (n \times n) \text{ Matrix}$$

$$\underline{b} = (n \times 1)\text{-Spaltenvektor}$$

$$\underline{c}^T = (1 \times n)\text{-Zeilenvektor}$$

$$\underline{x}(k) = (n \times 1)\text{-Zustandsvektor}$$

Bei einer unmittelbaren Realisierung nach diesen Beziehungen wird in jedem Filterschritt ein neuer Zustandsvektor $\underline{x}(k)$ berechnet. Dies ist aber im Prinzip nicht unbedingt erforderlich, denn aus einem bekannten $\underline{x}(k)$ lassen sich beliebig viele nachfolgende Ausgangswerte berechnen. Durch fortgesetztes Einsetzen von (6-1a,b) erhält man

$$y(k) = \underline{c}^T \underline{x}(k) + d v(k)$$

$$y(k+1) = \underline{c}^T \underline{A} \underline{x}(k) + \underline{c}^T \underline{b} v(k) + d v(k+1)$$

$$y(k+2) = \underline{c}^T \underline{A}^2 \underline{x}(k) + \underline{c}^T \underline{A} \underline{b} v(k) + \underline{c}^T \underline{b} v(k+1) + d v(k+2)$$

$$\vdots$$

$$y(k+\kappa) = \underline{c}^T \underline{A}^\kappa \underline{x}(k) + \sum_{i=0}^{\kappa-1} \underline{c}^T \underline{A}^{\kappa-i-1} \underline{b} v(k+i) + d v(k+\kappa)$$

(6-2)

Dies bildet die Grundlage für den z.B. in /40/ vorgeschlagenen Blockalgorithmus. Hierbei wird nur jeder p-te Zustandsvektor $\underline{x}(k)$, $\underline{x}(k+p)$, $\underline{x}(k+2p)$ usw. berechnet und die dazwischen liegenden

den Ausgangswerte jeweils nach den Beziehungen (6-2) ermittelt.
Zunächst gilt unter Beachtung von (6-1a)

$$\underline{x}(k+p) = \underline{A}^p \underline{x}(k) + \sum_{i=0}^{p-1} \underline{A}^{p-i-1} \underline{b} v(k+i) . \quad (6-3)$$

Zusammen mit (6-2) ergibt sich nun die Matrixbeziehung

$$\begin{matrix} n \\ \left\{ \begin{array}{c} \underline{x}(k+p) \\ \underline{y}(k) \\ \underline{y}(k+1) \\ \underline{y}(k+2) \\ \vdots \\ \underline{y}(k+p-1) \end{array} \right\} \\ p \end{matrix} = \begin{matrix} \overbrace{\begin{array}{c} \underline{A}^p \\ \underline{c}^T \underline{A} \\ \underline{c}^T \underline{A}^2 \\ \vdots \\ \underline{c}^T \underline{A}^{p-1} \end{array}}^n \quad \underbrace{\begin{array}{c} \underline{A}^{p-1} \underline{b} \quad \underline{A}^{p-2} \underline{b} \quad \underline{A}^{p-3} \underline{b} \quad \dots \quad \underline{b} \\ d \quad 0 \quad \dots \quad 0 \\ \underline{c}^T \underline{b} \quad d \quad 0 \quad \dots \quad 0 \\ \underline{c}^T \underline{A} \underline{b} \quad \underline{c}^T \underline{b} \quad d \quad 0 \quad \dots \quad 0 \\ \vdots \\ \underline{c}^T \underline{A}^{p-2} \underline{b} \quad \dots \quad \dots \quad d \end{array}}^p \end{matrix} \begin{matrix} \underline{x}(k) \\ \underline{v}(k) \\ \underline{v}(k+1) \\ \underline{v}(k+2) \\ \vdots \\ \underline{v}(k+p-1) \end{matrix} \quad (6-4)$$

Zur Bestimmung von p y-Werten einschließlich des neuen ZV-Vektors $\underline{x}(k+p)$ sind also jeweils (n+p) innere Produkte zu berechnen. Die Zahl der Multiplikationen pro Filterausgangswert ist demnach

$$z_M = \frac{1}{p} (n+p)^2 \quad (6-5)$$

Das Minimum wird für p = n erreicht, wobei dann

$$z_M = 4n \quad (6-6)$$

gilt. Zum Vergleich wird der Aufwand bei einigen konventionellen Filtern herangezogen:

a) Parallelanordnung mit Blöcken zweiten Grades in 1. kanonischer Form: $z_M = \frac{5}{2} n+1$

b) Parallelanordnung von Blöcken zweiten Grades nach Mullis/Roberts oder Barnes: $z_M = \frac{9}{2} n$

c) nicht blockstrukturiertes Normalform-Filter nach Barnes:

$$Z_M = (n+1)^2$$

Es zeigt sich, daß gerade die interessanten überlaufstabilen und rauscharmen Strukturen nach Mullis/Roberts und Barnes aufwendiger sind als das o.a. Verfahren, welches darüber hinaus noch die folgenden Vorteile bietet:

1. Die Berechnungen in (6-4) sind sehr gut parallelisierbar.
2. Die Methode eignet sich sehr gut für Rechner, die für die Auswertung von inneren Produkten mit einem Vektorprozessor ausgerüstet sind.
3. Bei Reduktion der Abtastrate am Ausgang, müssen nur die interessierenden y -Werte berechnet werden.

6.3 Quantisierungseigenschaften

6.3.1 Rundungsrauschen

Die folgenden theoretischen Ergebnisse sind aus /40/ bzw. /42/ entnommen, wo sie ausführlich hergeleitet wurden. Sie werden hier der Vollständigkeit halber angegeben, wobei aber auf einen Beweis verzichtet wird. Stattdessen soll eine anschauliche Interpretation, ergänzt durch Meßergebnisse, gegeben werden.

Das in (6-4) beschriebene Filter läßt sich als System mit p Eingängen $v_i(k) = v(kp+i)$, $i = 0, \dots, p-1$, und p Ausgängen $y_i(k) = y(kp+i)$ verstehen. Bezeichnet nun $\sigma_{y_i}^2$ die Varianz des Rundungsfehlers am Ausgang y_i , so gilt bei Rundung nach Zusammenfassen der Teilprodukte

$$\sigma_{y_i}^2 = \sigma_o^2 + \sigma_{y_i'}^2; \quad \sigma_{y_i'}^2 = \sigma_o^2 \sum_{v=0}^{\infty} \underline{c}^T \underline{A}^{vp+i} (\underline{A}^T)^{vp+i} \underline{c},$$

$$\sigma_o^2 = \frac{Q^2}{12} \quad (6-7)$$

wobei Q gleich dem Quantisierungsintervall ist. Der Term $\sigma_{y_i}^2$ kennzeichnet hier den durch Rundung der Zustandsvariablen hervorgerufenen Störanteil. Er wird in der Regel dominieren. $\sigma_{y_i}^2$ hängt von i ab, d.h. daß die Störleistung an den einzelnen Ausgängen unterschiedlich ist. Dies wird durch folgende Überlegung einsichtig: Jeder Fehler bei Rundung einer Zustandsvariablen erzeugt eine Impulsantwort von der Rundungsstelle zu den Ausgängen y_i , die mit wachsendem i abklingt. Demnach wird auch die Störleistung mit wachsendem i abnehmen (siehe Meßergebnisse in Abschnitt 6.4).

Aus (6-7) läßt sich nun die mittlere Störleistung aller Ausgänge zu /42/

$$\overline{\sigma_{y_i}^2} = \frac{1}{P} \sum_{i=0}^{P-1} \sigma_{y_i}^2 = \sigma_o^2 + \underbrace{\sigma_o^2 \frac{1}{P} \sum_{v=0}^{\infty} \underline{c}^T \underline{A}^v (\underline{A}^T)^v \underline{c}}_{:= \sigma_{y_i}^2} \quad (6-8)$$

bestimmen. Diese ist offenbar gleich der Rauschleistung der Ausgangsfolge $y(k)$. Wir vergleichen diesen Ausdruck nun mit der Störleistung σ^2 am Ausgang einer konventionellen Zustandsraum-Struktur. Nach Anhang A3, Glg. (A3-33) gilt

$$\sigma^2 = \sigma_o^2 + \sigma_o^2 \text{ Spur } \{\underline{W}\} , \quad (6-9a)$$

wobei $\underline{W}$ gemäß (A3-31) als

$$\underline{W} = \sum_{v=0}^{\infty} (\underline{c}^T \underline{A}^v)^T \underline{c}^T \underline{A}^v \quad (6-9b)$$

gegeben ist. Damit folgt

$$\begin{aligned} \text{Spur}\{\underline{W}\} &= \sum_{v=0}^{\infty} \text{Spur} \{ (\underline{c}^T \underline{A}^v)^T (\underline{c}^T \underline{A}^v) \} = \\ &= \sum_{v=0}^{\infty} \text{Spur} \{ (\underline{c}^T \underline{A}^v) \cdot (\underline{c}^T \underline{A}^v)^T \} = \\ &= \sum_{v=0}^{\infty} \text{Spur} \{ \underline{c}^T \underline{A}^v (\underline{A}^T)^v \underline{c} \} = \sum_{v=0}^{\infty} \underline{c}^T \underline{A}^v (\underline{A}^T)^v \underline{c} \end{aligned}$$

und mit (6-9a) erhält man

$$\sigma^2 = \sigma_o^2 + \underbrace{\sigma_o^2 \sum_{v=0}^{\infty} \underline{c}^T \underline{A}^v (\underline{A}^T)^v \underline{c}}_{= \sigma'^2} \quad (6-9c)$$

σ'^2 beschreibt auch hier den durch Rundung der Zustandsgrößen entstehenden Störanteil. Ein Vergleich mit (6-8) liefert nun den Zusammenhang

$$\overline{\sigma_{y_i}^2} = \frac{1}{p} \sigma'^2. \quad (6-10)$$

Der dominierende Störanteil wird also um den Faktor $\frac{1}{p}$ in seiner Leistung reduziert. Auch dies läßt sich anschaulich verstehen: Bei Blockverarbeitung wird nur jeder p-te Zustandsvektor berechnet. Im Intervall $|k, k+p-1|$ entsteht demnach nur einmal ein Rundungsfehler. Demgegenüber wird bei einer konventionellen Zustandsraum-Struktur der Zustandsvektor in jedem Filterschritt neu berechnet und gerundet. Da die Fehler als unkorreliert betrachtet werden dürfen, wird infolgedessen die Störleistung um den Faktor p größer sein.

6.3.2 Überlaufstabilität

Eine konventionelle Zustandsraum-Struktur ist nach Kap. 2 antwortstabil, wenn die Bedingung

$$\|\underline{A} \underline{x}\|_{\underline{G}} \leq \|\underline{x}\|_{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n \quad (2-22a)$$

erfüllt ist und die Überlaufbehandlung (z.B. Sättigungskennlinie) am Eingang der Zustandsspeicher erfolgt. Bei der Herleitung wurden allgemeine Systeme mit einer beliebigen Anzahl von Ein- und Ausgängen zugelassen. Demnach handelt es sich bei dem o.a. Blockverarbeitungssystem lediglich um einen Spezialfall. Die Systemmatrix ist hier allerdings $\underline{A}^p$. Man zeigt nun sehr einfach, daß

das Verfahren antwortstabil ist, wenn die $\underline{A}$ -Matrix der zugrundeliegenden Filterstruktur die Bedingung (2-22a) erfüllt. Da (2-22a) für alle $\underline{x} \in \mathbb{R}^n$ gelten muß, also auch für $\underline{x}_1 = \underline{A} \underline{x}$, folgt

$$||\underline{A} \underline{x}_1||_{\underline{G}} = ||\underline{A}^2 \underline{x}||_{\underline{G}} \leq ||\underline{x}_1||_{\underline{G}} = ||\underline{A} \underline{x}||_{\underline{G}} \leq ||\underline{x}||_{\underline{G}}$$

$$\text{bzw. } ||\underline{A}^2 \underline{x}||_{\underline{G}} \leq ||\underline{x}||_{\underline{G}}$$

und allgemein

$$||\underline{A}^p \underline{x}||_{\underline{G}} \leq ||\underline{x}||_{\underline{G}} \quad (6-12)$$

Daneben besteht grundsätzlich auch die Möglichkeit, daß $\underline{A}^p$ die Bedingung (6-12) erfüllt, obwohl $\underline{A}$ nicht die Eigenschaft (2-22a) besitzt. In solchen Fällen wird das Filter erst aufgrund der Blockverarbeitung antwortstabil.

Wir betrachten nun speziell solche $\underline{A}$ -Matrizen, die durch Transformation mit einer Diagonalmatrix normal gemacht werden können. Dann gilt

$$||\underline{A} \underline{x}||_{\underline{G}} \leq \max_v \{|\lambda_v|\} ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n, \quad (6-13)$$

wobei λ_v die Eigenwerte von $\underline{A}$ bezeichnet, die i.a. mit den Filterpolstellen identisch sind. Es folgt die (6-12) entsprechende Beziehung

$$||\underline{A}^p \underline{x}||_{\underline{G}} \leq \left[\max_v \{|\lambda_v|\} \right]^p ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n$$

bzw.

$$||\underline{A}^p \underline{x}||_{\underline{G}} \leq \max_v \{|\lambda_v^p|\} ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n, \quad (6-14)$$

wobei die λ_v^p , $v = 1(1)n$, gleich den Eigenwerten von $\underline{A}^p$ sind.

Schließlich sei noch vermerkt, daß $\underline{A}^p$ normal ist, sofern $\underline{A}$ eine normale Matrix ist. Zum Beweis schreiben wir $\underline{A}$ in der Form

$$\underline{A} = \underline{U} \underline{\Lambda} \underline{U}^{*T}$$

$$\underline{U} = \text{unitäre Modalmatrix mit } \underline{U}^{*T} \underline{U} = \underline{E}$$

$$\underline{\Lambda} = \text{Diagonalmatrix der Eigenwerte von } \underline{A}.$$

Dann ist

$$\underline{A}^p = \underline{U} \underline{\Lambda}^p \underline{U}^{*T}$$

und somit normal. Diese Eigenschaft wurde bereits von Barnes in den eingangs zitierten Arbeiten berücksichtigt.

6.3.3 Granulare Grenzzyklen

Wir betrachten insbesondere Basis-Strukturen mit der Eigenschaft (6-13). Dann gilt im Fall der Blockverarbeitung gemäß (6-14)

$$||\underline{A}^p \underline{x}||_{\underline{G}} \leq \max_v \{ |\lambda_v^p| \} ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \in \mathbb{R}^n.$$

Unter Berücksichtigung der Ergebnisse aus Abschnitt 5.4.2 erhält man bei Rundung folgende obere Schranke für die Zustandsvektoren von Grenzzyklen:

$$||\underline{x}||_{\underline{G}} \leq Q \frac{\frac{1}{2} \sqrt{\sum_{\mu=1}^n G_{\mu}}}{1 - \max_v \{ |\lambda_v^p| \}} \quad (6-15)$$

Diese Schranke verringert sich offenbar für wachsende Segmentlänge p.

Wenn die Basis-Struktur aus Blöcken zweiten Grades in Parallelanordnung besteht, so ist das System für $|\lambda_v^p| < \frac{1}{2}$, $v = 1(1)n$, grenzzykklusfrei. Dies ist z.B. bei einer Segmentlänge $p = 10$ bereits für Filterpolstellen $|\lambda_v| \leq 0,933$ der Fall.

6.4 Meßergebnisse

Der o.a. Filteralgorithmus wurde mit Hilfe des Mikrosignalprozessors MSP 82 /44/ realisiert. Die in diesem Zusammenhang notwendigen Programmierarbeiten wurden im Rahmen einer Diplomarbeit /45/ durchgeführt. Als Basisstruktur wurde eine Parallelanordnung von Filterblöcken zweiten Grades nach Mullis und Roberts verwendet, da diese sich durch Antwortstabilität und sehr geringes Rundungsrauschen auszeichnet. Die Rundung erfolgt nach Zusammenfassung der Teilprodukte.

Bild 6.1 zeigt gemessene Rauschzahlen R_{i_v} an den Ausgängen y_v in Abhängigkeit von der Durchlaßgrenzfrequenz eines Tiefpasses 10-ten Grades. Bild 6.2 zeigt den Einfluß der Segmentlänge p , wobei für festes p jeweils die Rauschzahlen R_{i_v} , $v = O(1)p-1$, aufgetragen sind.

Für die Messung der granularen Grenzzyklen wurden Filterblöcke zweiten Grades mit den Nullstellen $z_{o1,2} = e^{\pm j30^\circ}$ und den Polstellen $z_{\infty 1,2} = 0,99e^{\pm j\psi}$ herangezogen. Filterausgang und Zustandsgrößen sind gemäß der L_∞ -Norm skaliert.

In Tabelle 6.1 sind für unterschiedliche Polwinkel ψ und Segmentlängen p die gefundenen maximalen Grenzzyklus-Amplituden angegeben.

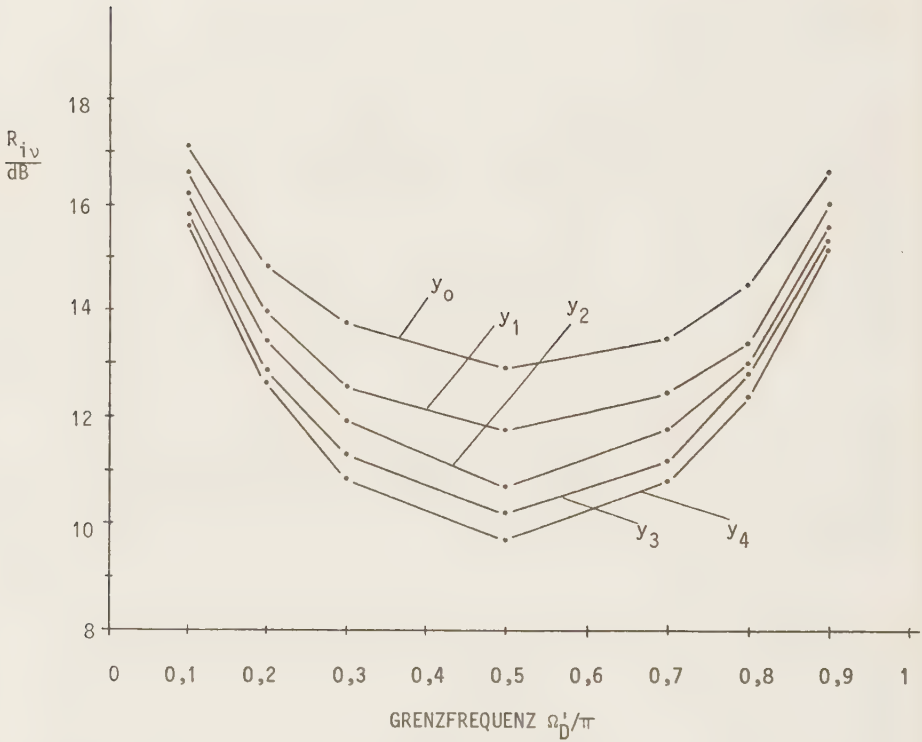


Bild 6.1: Gemessene Rauschzahlen der Ausgänge y_0 bis y_5 bei Tiefpässen 10-ten Grades der Durchlaßgrenzfrequenz Ω_D' .

Blocklänge: $p = 5$

Variation von Ω_D' durch Allpaßtransformation; Prototyp-Filter nach Tabelle 5.1. Skalierung von Zustandsvariablen und Filterausgang nach der L_∞ -Norm



Bild 6.2: Gemessene Rauschzahlen der Ausgänge y_0 bis y_{p-1} in Abhängigkeit von der Blocklänge p .
Filter: Tiefpaß 10ten Grades nach Tabelle 5.1, Skalierung nach der L_∞ -Norm

Segment- länge p	ρ^p	maximale Amplitude am Ausgang		
		$\psi = 1^\circ$	$\psi = 45^\circ$	$\psi = 90^\circ$
1	0,99	34 Q	52 Q	69 Q
2	0,9801	16 Q	16 Q	18 Q
4	0,9606	4 Q	8 Q	8 Q
6	0,9415	2 Q	5 Q	6 Q
8	0,9227	1 Q	4 Q	4 Q
10	0,9044	1 Q	3 Q	4 Q
12	0,8864	1 Q	3 Q	3 Q
14	0,8687	1 Q	2 Q	2 Q
16	0,8515	1 Q	2 Q	2 Q
18	0,8345	1 Q	2 Q	2 Q
20	0,8179	1 Q	1 Q	1 Q

Tabelle 6.1: Grenzyklen eines Filterblockes zweiten Grades
 $z_{01,2} = e^{\pm j30^\circ}$, $z_{\infty 1,2} = 0,99 e^{\pm j\psi}$

7. STABILITÄT ZEITVARIABLER DIGITALER FILTER

7.1 Vorbemerkungen

Bislang wurden ausschließlich zeitinvariante Systeme untersucht. Für die Analyse der Überlaufstabilität wurden hierbei u.a. Zustandsraum-Betrachtungen angestellt. Die dabei gewonnenen Erkenntnisse lassen sich aber, mit gewissen Einschränkungen, auch für die Beurteilung der Stabilität zeitvarianter Digitalfilter heranziehen. Daher soll in dem folgenden abschließenden Kapitel dieser Nebenaspekt angesprochen werden, wobei insbesondere der Einfluß der Filterstruktur herausgestellt wird. Die Ergebnisse wurden teilweise in /46/ dargelegt.

Wir betrachten ein lineares zeitvariables digitales System, das im homogenen Fall durch

$$\underline{x}(k+1) = \underline{A}(k) \underline{x}(k) \quad (7-1)$$

beschrieben sei. Die zeitabhängige Systemmatrix $\underline{A}(k)$ sei als

$$\underline{A}(k) = \underline{A}_\mu = \text{konst.}, \quad k_\mu \leq k < k_{\mu+1} \quad (7-2)$$

gegeben, wobei die Eigenwerte $z_{\omega_\nu}^{(\mu)}$ der $\underline{A}_\mu$ im Innern des Einheitskreises liegen sollen. Die Variation des Systems geschehe also dadurch, daß zwischen Koeffizientensätzen stabiler Systeme umgeschaltet wird. Zeitvariable Filter der beschriebenen Art können instabil sein, obwohl die Eigenwerte der Teilsysteme im Innern des Einheitskreises liegen (z.B. /47/).

Dieses Verhalten wird zunächst anschaulich erklärt und mit Beispielen illustriert. Auf die analytischen Zusammenhänge wird hingewiesen. Sodann wird gezeigt, wie durch Wahl geeigneter Filterstrukturen die Stabilität stets gesichert werden kann. Schließ-

lich wird eine Transformation angegeben, bei deren Anwendung das Systemverhalten unabhängig von der gewählten Filterstruktur ist.

7.2 Stabilität eines Filters zweiten Grades am Beispiel der ersten kanonischen Form

Für konjugiert komplexe Eigenwerte $z_{\infty 1,2}^{(\mu)} = \rho_{\mu} e^{\pm j\psi_{\mu}}$ gilt

$$\underline{A} = \begin{bmatrix} +2\rho_{\mu} \cos \psi_{\mu} & 1 \\ -\rho_{\mu}^2 & 0 \end{bmatrix} . \quad (7-3)$$

Im Intervall $k_{\mu} \leq k < k_{\mu+1}$ ist $\underline{A}$ konstant, und im homogenen Fall läßt sich die Zustandsvariable $x_1(k)$ stets in der Form

$$x_1(k) = \hat{x} \rho_{\mu}^k \cos(\psi_{\mu} k + \theta) \quad (7-4a)$$

schreiben, wobei die Konstanten $\hat{x}$ und θ durch den gegebenen Anfangszustand festgelegt sind. Wegen (7-3) folgt

$$\begin{aligned} x_2(k) &= -\rho_{\mu}^2 x_1(k-1) \\ &= -\hat{x} \rho_{\mu}^{k+1} \cos[\psi_{\mu}(k-1) + \theta] . \end{aligned} \quad (7-4b)$$

Im Grenzfall $\rho_{\mu} \rightarrow 1$ bewegt sich der Zustandsvektor also auf einer konzentrischen Ellipse, deren Form durch den Polwinkel ψ bestimmt wird.

In Bild 7.1 sind zwei Beispiele für $\rho = 1$ und unterschiedliche ψ angegeben.

Anhand dieser Beispiele läßt sich die mögliche Instabilität eines zeitvarianten Systems anschaulich erklären: Es sei zunächst angenommen, daß sich der Zustandsvektor nach Vorgabe eines Anfangszustandes auf der in Bild 7.2 gezeichneten Ellipse 1 bewegt. In

Punkt I werde der Koeffizientensatz umgeschaltet, so daß sich der Zustandsvektor im folgenden auf einer Ellipse gemäß Bild 7.1b bewegen muß. Nach Erreichen von Punkt II werde erneut auf den ersten Koeffizientensatz umgeschaltet usw. Die Gründe für das Anwachsen des ZV-Vektors sind offenbar:

1. Auch im zeitinvarianten Fall kann die Länge (euklidische Norm) des ZV-Vektors vorübergehend zunehmen.
2. Der ZV-Vektor bewegt sich auf Ellipsen, deren Form vom Polwinkel ψ abhängt.

Durch Wahl anderer Umschaltzeitpunkte kann offenbar aber auch ein Abklingen der Zustandsvariablen erreicht werden.

Die angestellten Überlegungen sind prinzipiell auf Filter mit Polstellen $[z_{\infty 1,2}] = \rho \neq 1$ übertragbar. Bild 7.3 zeigt hierzu das Ausschwingverhalten zweier stabiler zeitinvarianter Systeme. Die einzelnen Werte liegen auf elliptischen Spiralen, die, entsprechend den unterschiedlichen Parametern, unterschiedliche Form haben. Ein periodisches Umschalten zwischen den beiden Systemen führt, abhängig von der Wahl der Schaltzeitpunkte, zu stabilem oder instabilem Verhalten (Bild 7.4a,b). Die Größen $\Delta k_{1,2}$ kennzeichnen hier die Anzahl der Filterschritte zwischen den Schaltzeitpunkten.

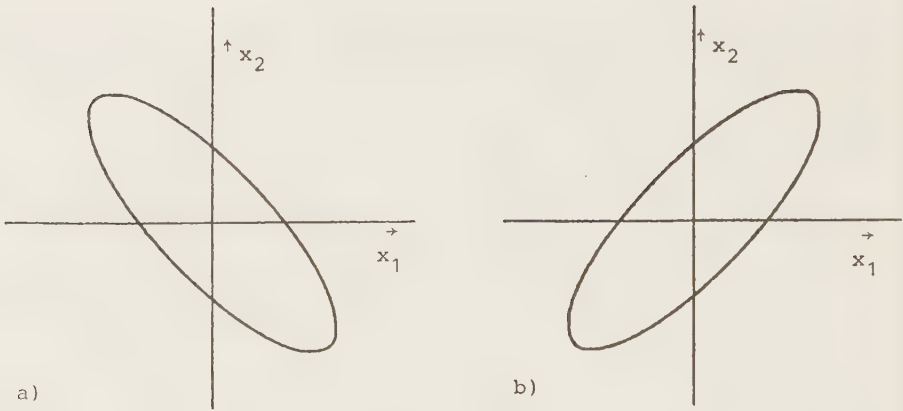


Bild 7.1: Ellipsen, auf die sich der Zustandsvektor für Polradius $\rho = 1$ bewegt
a) Polwinkel $\psi_1 = 20^\circ$
b) Polwinkel $\psi_2 = 160^\circ$

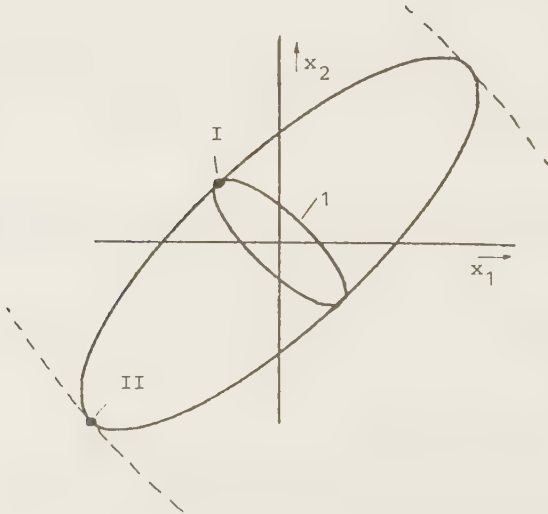


Bild 7.2: Zur Veranschaulichung des Stabilitätsverhaltens zeitvarianter Systeme

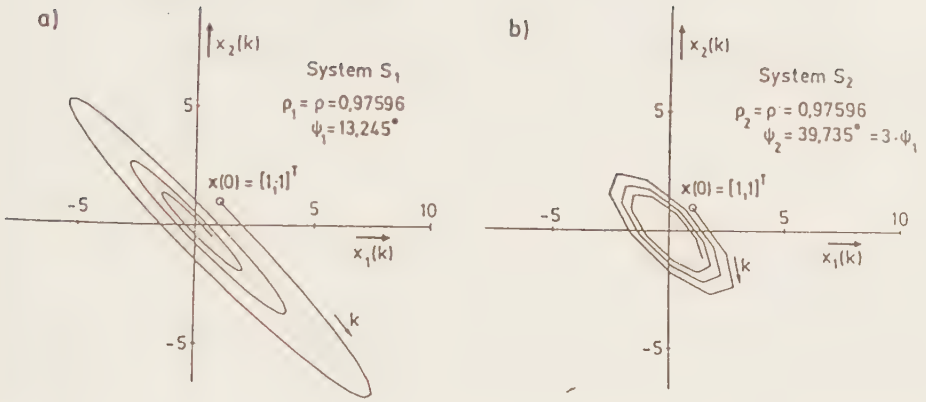


Bild 7.3: Ausschwingverhalten zweier stabiler Systeme

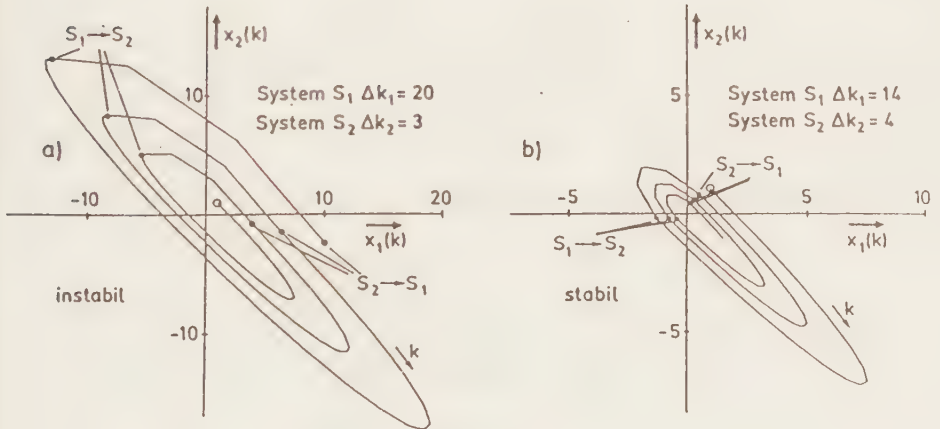


Bild 7.4: Zustandsvektoren bei Umschaltungen zwischen den Systemen S_1 und S_2

7.3 Analytische Beschreibung periodisch zeitvarianter Systeme

Bekanntlich ist die allgemeine Lösung von (7-1)

$$\underline{x}(k) = \begin{bmatrix} k-1 \\ \Pi \\ i=k_0 \end{bmatrix} \underline{A}(i) \underline{x}(k_0) = \underline{A}(k-1) \dots \underline{A}(k_0) \underline{x}(k_0) := \underline{\phi}(k, k_0) \cdot \underline{x}(k_0) \quad (7-5)$$

mit der Übergangsmatrix $\underline{\phi}(k, k_0)$ (z.B. /48/, /49/).

Wir nehmen jetzt an, daß (7-2) gilt und jede Matrix $\underline{A}_\mu$ für ein Intervall $\Delta k_\mu = k_{\mu+1} - k_\mu$ wirksam ist. Weiterhin möge die Umschaltung periodisch erfolgen mit der Periode K_0 . Es gelte also .

$$\underline{A}(k+K_0) = \underline{A}(k) \quad (7-6a)$$

mit

$$K_0 = \sum_{\mu=1}^{\ell} \Delta k_\mu \quad (7-6b)$$

Man sieht leicht ein, daß in diesem Fall neben $\underline{\phi}(k, k_0)$ auch

$$\underline{\phi}(k_0 + k + K_0, k_0) = \underline{\phi}(k + k_0, k_0) \cdot \underline{K} \quad (7-7a)$$

eine Lösung von (7-1) ist, wobei

$$\underline{K} = \underline{A}_{\ell}^{\Delta k_\ell} \underline{A}_{\ell-1}^{\Delta k_{\ell-1}} \dots \underline{A}_1^{\Delta k_1} = \underline{\phi}(k_0 + K_0, k_0) \quad (7-7b)$$

eine konstante Matrix ist. Offensichtlich ergibt sich aus (7-7a,b) für das Verhalten in der m-ten Periode

$$\underline{\phi}(k + k_0 + m \cdot K_0, k_0) = \underline{\phi}(k + k_0, k_0) \cdot \underline{K}^m \quad (7-8)$$

Da $\underline{\phi}(k + k_0, k_0)$ mit $k < K_0$ beschränkt ist, sind offenbar die Eigenschaften von $\underline{K}$ maßgebend für die Stabilität: Das System ist genau dann asymptotisch stabil, wenn alle Eigenwerte von $\underline{K}$ im Innern des Einheitskreises liegen. (Siehe /49, 5C, 52/).

Ergänzend sei noch vermerkt, daß die Stabilität nicht davon abhängt, mit welcher Systemmatrix $\underline{A}_\mu$ die Filterung beginnt, solange die Reihenfolge der $\underline{A}_\mu$ ansonsten erhalten bleibt. Wird z.B. anstelle von $\underline{A}_1$ mit $\underline{A}_\ell$ begonnen, so ergibt sich eine gegenüber (7-7b) geänderte $\underline{K}$ -Matrix

$$\underline{K}' = \underline{A}_{\ell-1}^{\Delta k_{\ell-1}} \dots \underline{A}_2^{\Delta k_2} \underline{A}_1^{\Delta k_1} \underline{A}_\ell^{\Delta k_\ell}. \quad (7-9)$$

Allgemein erhält man offenbar $\underline{K}'$ aus $\underline{K}$ durch zyklisches Vertauschen der $\underline{A}_i^{\Delta k_i}$, $i = 1(1)\ell$. Im allgemeinen ist dann zwar $\underline{K}' \neq \underline{K}$, doch bleiben die Eigenwerte und somit das Stabilitätsverhalten unverändert /20/. Um dies einzusehen schreiben wir $\underline{K}$ bzw. $\underline{K}'$ in der Form

$$\underline{K} = \underline{\phi}_{II} \cdot \underline{\phi}_I$$

bzw.

(7-10)

$$\underline{K}' = \underline{\phi}_I \cdot \underline{\phi}_{II}$$

mit

$$\begin{aligned} \underline{\phi}_I &= \underline{A}_{\ell 1}^{\Delta k_{\ell 1}} \underline{A}_{\ell 1-1}^{\Delta k_{\ell 1-1}} \dots \underline{A}_1^{\Delta k_1} \\ \underline{\phi}_{II} &= \underline{A}_\ell^{\Delta k_\ell} \underline{A}_{\ell-1}^{\Delta k_{\ell-1}} \dots \underline{A}_{\ell_1+1}^{\Delta k_{\ell_1+1}}, \quad \ell_1 < \ell. \end{aligned}$$

Sei nun λ ein Eigenwert von $\underline{K}$ und $\underline{x}$ der zugehörige Eigenvektor, so gilt

$$\underline{K} \underline{x} = \lambda \underline{x} \quad \text{bzw.} \quad \underline{\phi}_{II} \underline{\phi}_I \underline{x} = \lambda \underline{x}.$$

Durch linksseitige Multiplikation mit $\underline{\phi}_I$ gewinnt man

$$\begin{aligned} \underbrace{\underline{\phi}_I \underline{\phi}_{II}}_{= \underline{K}'} \underbrace{\underline{\phi}_I \underline{x}}_{= \underline{x}'} &= \underbrace{\lambda \underline{\phi}_I \underline{x}}_{= \underline{x}'} \\ &= \underline{K}' \underline{x}' = \underline{x}' \end{aligned}$$

λ ist folglich auch Eigenwert von $\underline{K}'$ mit dem Eigenvektor

$$\underline{x}' = \underline{\phi}_I \underline{x}.$$

7.4 Stabilität durch Wahl einer geeigneten Filterstruktur

Wie die Betrachtungen in Abschnitt 7.2 zeigten, kann die Länge des Zustandsvektors eines stabilen zeitinvarianten Filters unter gewissen Bedingungen vorübergehend größer werden. Dies ist die wesentliche Voraussetzung für die Instabilität im zeitvarianten Fall. Umgekehrt ist das System folglich stabil, wenn

$$||\underline{x}(k+1)|| < ||\underline{x}(k)|| \quad \forall k \text{ mit } \underline{x}(k) \neq \underline{0} \quad (7-11)$$

$$||\dots|| = \text{euklidische Norm}$$

bzw.

$$||\underline{A}(k) \underline{x}(k)|| < ||\underline{x}(k)|| \quad (7-12)$$

gilt (hinreichende Bedingung). (7-12) wird z.B. erfüllt, wenn $\underline{A}(k)$ eine normale Matrix ist, deren Eigenwerte im Innern des Einheitskreises liegen (siehe Normalform-Strukturen, Abschnitt 3.2.2).

Als Beispiel betrachten wir ein System zweiter Ordnung, dessen zeitvariable $\underline{A}$ -Matrix durch (7-2) beschrieben werde. Im Fall der Normalform-Struktur gilt dann

$$\underline{A}_\mu = \rho_\mu \begin{bmatrix} \cos \psi_\mu & \sin \psi_\mu \\ -\sin \psi_\mu & \cos \psi_\mu \end{bmatrix} \quad (7-13)$$

$$\text{Eigenwerte: } z_{\infty 1,2}^{(\mu)} = \rho_\mu e^{\pm j \psi_\mu}$$

und es folgt

$$||\underline{x}(k+1)|| = ||\underline{A}_\mu \underline{x}(k)|| = \rho_\mu ||\underline{x}(k)||; k_\mu \leq k < k_{\mu+1} \quad (7-14)$$

Für $\rho_\mu < 1$ nimmt die Länge des ZV-Vektors $\underline{x}(k)$ in jedem Filterschritt ab. Die hiermit allgemein gewährleistete Stabilität gilt

Eine weitere Filterstruktur ist die "Normalized ladder" Form von Markel und Gray (siehe Abschnitt 3.4). Hier gilt

$$||\underline{A}(k) \underline{x}(k)|| \leq ||\underline{x}(k)|| \quad \forall k \quad (7-17)$$

Im Gegensatz zur Forderung (7-12) ist hier das Gleichheitszeichen für spezielle $\underline{x}(k) \neq \underline{0}$ zwar prinzipiell möglich, was aber praktisch keine wesentliche Einschränkung darstellt.

7.5 Äquivalente Realisierungen mit unterschiedlichen Filterstrukturen

Die bisherigen Überlegungen haben gezeigt, daß die Systemreaktion eines zeitvariablen Filters von der jeweiligen Struktur abhängt. Diese Abhängigkeit läßt sich aber durch geeignete Maßnahmen beseitigen, wie im folgenden erläutert wird.

Wir kehren zunächst noch einmal zum Fall des zeitinvarianten Filters zurück. Hier läßt sich der n-dimensionale Zustandsvektor $\underline{x}(k)$ mit Hilfe einer nichtsingulären konstanten $(n \times n)$ Matrix $\underline{T}$ in den modifizierten Zustandsvektor $\underline{x}'(k) = \underline{T}^{-1} \underline{x}(k)$ transformieren (siehe Abschnitt 2.2.1). Die ursprüngliche Zustandsraum-Darstellung mit den Systemmatrizen $\underline{A}$, $\underline{B}$, $\underline{C}$, $\underline{D}$ geht dabei über in

$$\begin{aligned} \underline{x}'(k+1) &= \underline{T}^{-1} \underline{A} \underline{T} \underline{x}'(k) + \underline{T}^{-1} \underline{B} \underline{v}(k) \\ \underline{y}(k) &= \underline{C} \underline{T} \underline{x}'(k) + \underline{D} \underline{v}(k) \end{aligned} \quad (7-18)$$

Entsprechend gehen wir nun im zeitvarianten Fall vor, wobei

$$\begin{aligned} \underline{A}(k) &= \underline{A}_\mu, \quad \underline{B}(k) = \underline{B}_\mu \\ \underline{C}(k) &= \underline{C}_\mu, \quad \underline{D}(k) = \underline{D}_\mu \end{aligned} \quad ; \quad k_\mu \leq k < k_{\mu+1} \quad (7-19)$$

gelte. Ferner sei

$$\underline{x}'(k) = \underline{T}_{\mu}^{-1} \underline{x}(k) \quad ; \quad k_{\mu} \leq k < k_{\mu+1} \quad , \quad (7-20)$$

wobei i.a. $\underline{T}_{\mu} \neq \underline{T}_{\nu}$ für $\mu \neq \nu$ ist. Daraus gewinnt man die modifizierten Zustandsraum-Darstellungen

$$\underline{x}'(k+1) = \underline{T}_{\mu}^{-1} \underline{A}_{\mu} \underline{T}_{\mu} \underline{x}'(k) + \underline{T}_{\mu}^{-1} \underline{B}_{\mu} \underline{v}(k) \quad (7-21a)$$

$$\underline{y}(k) = \underline{C}_{\mu} \underline{T}_{\mu} \underline{x}'(k) + \underline{D}_{\mu} \underline{v}(k)$$

$$\text{für } k_{\mu} \leq k < k_{\mu+1} - 1$$

und für $k = k_{\mu+1} - 1$

$$\underline{x}'(k+1) = \underline{T}_{\mu+1}^{-1} \underline{A}_{\mu} \underline{T}_{\mu} \underline{x}'(k) + \underline{T}_{\mu+1}^{-1} \underline{B}_{\mu} \underline{v}(k) \quad (7-21b)$$

$$\underline{y}(k) = \underline{C}_{\mu} \underline{T}_{\mu} \underline{x}'(k) + \underline{D}_{\mu} \underline{v}(k)$$

Für die modifizierten Systemmatrizen folgt mit (7-21a,b)

$$\underline{A}'(k) = \underline{T}_{\mu}^{-1} \underline{A}_{\mu} \underline{T}_{\mu} := \underline{A}'_{\mu}$$

$$\underline{B}'(k) = \underline{T}_{\mu}^{-1} \underline{B}_{\mu} := \underline{B}'_{\mu}$$

$$\text{für } k_{\mu} \leq k < k_{\mu+1} - 1$$

$$\underline{C}'(k) = \underline{C}_{\mu} \underline{T}_{\mu} := \underline{C}'_{\mu}$$

$$\underline{D}'(k) = \underline{D}_{\mu} := \underline{D}'_{\mu} \quad (7-22a)$$

und für $k = k_{\mu+1} - 1$

$$\underline{A}'(k) = \underline{T}_{\mu+1}^{-1} \underline{A}_{\mu} \underline{T}_{\mu} = \underline{T}_{\mu+1}^{-1} \underline{T}_{\mu} \underline{A}'_{\mu}$$

$$\underline{B}'(k) = \underline{T}_{\mu+1}^{-1} \underline{B}_{\mu} = \underline{T}_{\mu+1}^{-1} \underline{T}_{\mu} \underline{B}'_{\mu}$$

$$\underline{C}'(k) = \underline{C}_{\mu} \underline{T}_{\mu} = \underline{C}'_{\mu}$$

$$\underline{D}'(k) = \underline{D}_{\mu} = \underline{D}'_{\mu} \quad (7-22b)$$

Dieses Ergebnis besagt, daß bei einer Transformation entsprechend (7-22a,b) die Systemeigenschaften nach außen hin unverändert blei-

ben. Der Unterschied zum zeitinvarianten Fall besteht lediglich in einer zusätzlichen Modifikation der Matrizen $\underline{A}'_{\mu}$ und $\underline{B}'_{\mu}$ zum Zeitpunkt $k = k_{\mu+1} - 1$, was einer Umrechnung des Zustandsvektors $\underline{x}'(k)$ beim Umschalten des Koeffizientensatzes gleichkommt.

Ist nun die durch (7-19) beschriebene Bezugsstruktur stabil, so gilt dies natürlich auch für das transformierte System. Wie diese Stabilität im konkreten Fall erreicht wird, soll an einem einfachen Beispiel demonstriert werden. Wir betrachten ein Filter zweiten Grades in erster kanonischer Form mit der in (7-3) gegebenen A-Matrix. Das System soll stabil sein, sofern die Polstellen der zeitinvarianten Teilsysteme im Innern des Einheitskreises liegen. Wir wählen nun eine stabile Bezugsstruktur gemäß (7-13). Für

$$\underline{A}_{\mu} = \rho_{\mu} \begin{bmatrix} \cos \psi_{\mu} & \sin \psi_{\mu} \\ -\sin \psi_{\mu} & \cos \psi_{\mu} \end{bmatrix}$$

$$\underline{A}'_{\mu} = \begin{bmatrix} 2\rho_{\mu} \cos \psi_{\mu} & 1 \\ -\rho_{\mu}^2 & 0 \end{bmatrix}$$

findet man eine mögliche Transformationsmatrix

$$\underline{T}_{\mu} = \begin{bmatrix} \rho_{\mu} \sin \psi_{\mu} & 0 \\ \rho_{\mu} \cos \psi_{\mu} & 1 \end{bmatrix},$$

so daß $\underline{A}'_{\mu} = \underline{T}_{\mu}^{-1} \underline{A}_{\mu} \underline{T}_{\mu}$ erfüllt ist.

Das System ist stabil, wenn

$$\underline{A}'(k) = \underline{A}'_{\mu} \quad \text{für} \quad k_{\mu} \leq k < k_{\mu+1} - 1$$

und

$$\underline{A}'(k) = \underline{T}_{\mu+1}^{-1} \underline{T}_{\mu} \underline{A}'_{\mu} \quad \text{für} \quad k = k_{\mu+1} - 1$$

gesetzt wird.

A N H A N G A 1

Erläuterungen zur Stabilitätsbedingung (2-16a,b)

Für ein Filter zweiten Grades soll der Ausdruck

$$\frac{||\underline{A} \underline{x}||_{\underline{G}}}{||\underline{x}||_{\underline{G}}} = \frac{\sqrt{\underline{x}^T \underline{A}^T \underline{G} \underline{A} \underline{x}}}{\sqrt{\underline{x}^T \underline{G} \underline{x}}} \quad (A1-1)$$

mit

$$\underline{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}; \quad \underline{G} = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}, \quad G_{1,2} > 0; \quad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

untersucht werden.

Wegen

$$\begin{aligned} ||\underline{A} \underline{x}||_{\underline{G}}^2 &= \underline{x}^T \underline{A}^T \underline{G} \underline{A} \underline{x} = \underline{x}^T \underline{G}^{1/2} \underline{G}^{-1/2} \underline{A}^T \underline{G}^{1/2} \underbrace{\underline{G}^{1/2} \underline{A} \underline{G}^{-1/2}}_{:= \hat{\underline{A}}} \underbrace{\underline{G}^{1/2} \underline{x}}_{:= \hat{\underline{x}}} \\ &= \hat{\underline{x}}^T \hat{\underline{A}}^T \hat{\underline{A}} \hat{\underline{x}} = ||\hat{\underline{A}} \hat{\underline{x}}||_{\underline{E}}^2 := ||\hat{\underline{A}} \hat{\underline{x}}||^2 \end{aligned} \quad (A1-2)$$

und

$$||\underline{x}||_{\underline{G}}^2 = \underline{x}^T \underline{G} \underline{x} = \hat{\underline{x}}^T \hat{\underline{x}} = ||\hat{\underline{x}}||_{\underline{E}}^2 = ||\hat{\underline{x}}||^2$$

folgt

$$\frac{||\underline{A} \underline{x}||_{\underline{G}}}{||\underline{x}||_{\underline{G}}} = \frac{||\hat{\underline{A}} \hat{\underline{x}}||}{||\hat{\underline{x}}||} \quad (A1-3)$$

Bei Verwendung von $\hat{\underline{A}}$ und $\hat{\underline{x}}$ anstelle von $\underline{A}$ und $\underline{x}$ kann die euklidische Vektornorm mit $\underline{G} = \underline{E}$ herangezogen werden. $\hat{\underline{A}}$ geht durch Ähnlichkeitstransformation aus $\underline{A}$ hervor und besitzt daher die gleichen Eigenwerte wie $\underline{A}$.

$$\hat{\underline{A}} = \begin{bmatrix} A_{11} & \sqrt{\frac{G_1}{G_2}} A_{12} \\ \sqrt{\frac{G_2}{G_1}} A_{21} & A_{22} \end{bmatrix} \quad (\text{A1-4})$$

Fall 1: $A_{12}A_{21} > 0$ (siehe (2-16a))

Wir wählen

$$\frac{G_1}{G_2} = \frac{A_{21}}{A_{12}} > 0 \quad \text{bzw.} \quad \sqrt{\frac{G_1}{G_2}} = \sqrt{\frac{A_{21}}{A_{12}}} \quad (\text{A1-5})$$

Dann ist

$$\hat{\underline{A}} = \begin{bmatrix} A_{11} & \text{sgn}\{A_{12}\} \cdot \sqrt{A_{12}A_{21}} \\ \text{sgn}\{A_{12}\} \cdot \sqrt{A_{12}A_{21}} & A_{22} \end{bmatrix}$$

eine symmetrische und somit normale Matrix /20/. Damit folgt (siehe auch Abschnitt 3.2.2)

$$\frac{||\hat{\underline{A}} \underline{\hat{x}}||}{||\underline{\hat{x}}||} \leq \max_v \{|\lambda_v|\} \quad , \quad (\text{A1-6})$$

wobei λ_v , $v = 1(1)n$ die Eigenwerte von $\hat{\underline{A}}$ bzw. $\underline{A}$ sind. Bei Filterstabilität (linearer Fall) gilt $|\lambda_v| < 1$ und es folgt

$$\frac{||\underline{A} \underline{x}||_{\underline{G}}}{||\underline{x}||_{\underline{G}}} = \frac{||\hat{\underline{A}} \underline{\hat{x}}||}{||\underline{\hat{x}}||} \leq \max_v \{|\lambda_v|\} < 1 \quad \forall \underline{x} \neq \underline{0} \quad (\text{A1-7})$$

bzw.

$$||\underline{A} \underline{x}||_{\underline{G}} < ||\underline{x}||_{\underline{G}} \quad \forall \underline{x} \neq \underline{0}$$

Fall 2: $A_{12}A_{21} < 0$ (siehe (2-16b))

Wir wählen nun

$$\frac{G_1}{G_2} = -\frac{A_{21}}{A_{12}} > 0 \quad (\text{A1-8})$$

und erhalten

$$\underline{\tilde{A}} = \underbrace{\begin{bmatrix} \operatorname{sgn}\{A_{12}\} & 0 \\ 0 & -\operatorname{sgn}\{A_{12}\} \end{bmatrix}}_{:= \underline{U}} \underbrace{\begin{bmatrix} \operatorname{sgn}\{A_{12}\} \cdot A_{11} & \sqrt{-A_{12}A_{21}} \\ \sqrt{-A_{12}A_{21}} & -\operatorname{sgn}\{A_{12}\}A_{22} \end{bmatrix}}_{:= \underline{\tilde{A}}} \quad (A1-9)$$

$\underline{\tilde{A}}$ ist symmetrisch und somit normal, während $\underline{U}$ eine Orthogonalmatrix darstellt. Damit folgt

$$||\underline{\tilde{A}} \underline{\hat{x}}|| = ||\underline{U} \underline{\tilde{A}} \underline{\hat{x}}|| = ||\underline{\tilde{A}} \underline{\hat{x}}|| \leq \max_v \{|\tilde{\lambda}_v|\} \cdot ||\underline{\hat{x}}|| \quad \forall \underline{\hat{x}} \neq \underline{0} \quad (A1-10)$$

$\tilde{\lambda}_v$, $v = 1, 2$, bezeichnen hier die Eigenwerte von $\underline{\tilde{A}}$. Die Stabilitätsbedingung $||\underline{\tilde{A}} \underline{\hat{x}}|| < ||\underline{\hat{x}}||$ ist also dann erfüllt, wenn $\max\{|\tilde{\lambda}_v|\} < 1$ gilt. Dazu betrachten wir das charakteristische Polynom der Matrix $\underline{\tilde{A}}$:

$$\begin{aligned} & (\tilde{\lambda} - \operatorname{sgn}\{A_{12}\}A_{11}) (\tilde{\lambda} + \operatorname{sgn}\{A_{12}\}A_{22}) + A_{12} A_{21} = \\ & = \tilde{\lambda}^2 + \tilde{\lambda} \operatorname{sgn}\{A_{12}\} [A_{22} - A_{11}] - [A_{11}A_{22} - A_{12}A_{21}] = \\ & = \tilde{\lambda}^2 + \tilde{\lambda} \operatorname{sgn}\{A_{12}\} [A_{22} - A_{11}] - \det[\underline{A}] := \tilde{\lambda}^2 + c_1 \tilde{\lambda} + c_0 \end{aligned} \quad (A1-11)$$

Bekanntlich gilt $|\tilde{\lambda}| < 1$, wenn die Bedingungen

$$|c_0| < 1 \quad \text{bzw.} \quad |\det [\underline{A}]| < 1 \quad (A1-12a)$$

und

$$|c_1| - c_0 < 1 \quad \text{bzw.} \quad |A_{11} - A_{22}| + \det [\underline{A}] < 1 \quad (A1-12b)$$

erfüllt sind, wobei die Gültigkeit von (A1-12a) bereits durch die Stabilität des Filters im linearen Fall gewährleistet ist. Damit ist (2-16b) bewiesen.

Im Grenzfall

$$|A_{11} - A_{22}| + \det [\underline{A}] = 1 \quad (A1-13)$$

folgt mit $\max \{|\tilde{\lambda}_v|\} = 1$

$$||\hat{\underline{A}} \hat{\underline{x}}|| \leq ||\hat{\underline{x}}|| \quad \forall \hat{\underline{x}}, \quad (\text{A1-14})$$

wobei das Gleichheitszeichen für spezielle $\hat{\underline{x}} \neq 0$ möglich ist.

Fall 3: $A_{12} A_{21} = 0$ mit $A_{12} = A_{21} = 0$ (siehe (2-16a))

Hier ist $\underline{A}$ eine Diagonalmatrix und somit normal. Es folgt unmittelbar

$$||\underline{A} \underline{x}|| \leq \max_v \{|\lambda_v|\} \cdot ||\underline{x}|| \quad \forall \underline{x} \neq 0 \quad (\text{A1-15})$$

Fall 4: $A_{12} A_{21} = 0$ mit $A_{21} = 0$, $A_{12} \neq 0$

$$\hat{\underline{A}} = \begin{bmatrix} A_{11} & \sqrt{\frac{G_1}{G_2}} A_{12} \\ 0 & A_{22} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \sqrt{\frac{G_1}{G_2}} A_{12} \\ 0 & \lambda_2 \end{bmatrix} \quad (\text{A1-16})$$

$\lambda_{1,2}$ sind die reellen Eigenwerte von $\underline{A}$ bzw. $\hat{\underline{A}}$. Wir betrachten den Ausdruck

$$||\hat{\underline{A}} \hat{\underline{x}}||^2 = \lambda_1^2 \hat{x}_1^2 + 2\lambda_1 \sqrt{\frac{G_1}{G_2}} A_{12} \hat{x}_1 \hat{x}_2 + \frac{G_1}{G_2} A_{12}^2 \hat{x}_2^2 + \lambda_2^2 \hat{x}_2^2 \quad (\text{A1-17})$$

Wegen

$$2\lambda_1 \sqrt{\frac{G_1}{G_2}} A_{12} \hat{x}_1 \hat{x}_2 \leq |\lambda_1 \cdot \sqrt{\frac{G_1}{G_2}} A_{12}| \cdot [\hat{x}_1^2 + \hat{x}_2^2]$$

folgt mit $|\lambda|_{\max} = \max\{|\lambda_{1,2}|\}$

$$||\hat{\underline{A}} \hat{\underline{x}}||^2 \leq \hat{x}_1^2 \left[\lambda_1^2 + |\lambda_1| \sqrt{\frac{G_1}{G_2}} |A_{12}| \right] + \hat{x}_2^2 \left[\lambda_2^2 + |\lambda_1| \sqrt{\frac{G_1}{G_2}} |A_{12}| + \frac{G_1}{G_2} A_{12}^2 \right] < \left[|\lambda|_{\max}^2 + |\lambda|_{\max} \sqrt{\frac{G_1}{G_2}} |A_{12}| + \frac{G_1}{G_2} A_{12}^2 \right] [\hat{x}_1^2 + \hat{x}_2^2] \quad (\text{A1-18})$$

Wegen $|\lambda|_{\max}^2 < 1$ läßt sich durch hinreichend kleines G_1/G_2 stets erreichen, daß

$$||\hat{\underline{A}} \hat{\underline{x}}||^2 < \hat{x}_1^2 + \hat{x}_2^2 = ||\hat{\underline{x}}||^2 \quad \text{bzw.} \quad ||\hat{\underline{A}} \hat{\underline{x}}|| < ||\hat{\underline{x}}|| \quad (\text{A1-19})$$

erfüllt ist. Im Grenzfall $G_1/G_2 \rightarrow 0$ gilt

$$||\hat{\underline{A}} \hat{\underline{x}}|| \leq |\lambda|_{\max} ||\hat{\underline{x}}||. \quad (\text{A1-20})$$

Dieses für normale Matrizen $\hat{\underline{A}}$ typische Ergebnis ist nicht erstaunlich, da $\hat{\underline{A}}$ im erwähnten Grenzfall zur Diagonalmatrix wird.

Fall 5: $A_{12}A_{21} = 0$ mit $A_{12} = 0, A_{21} \neq 0$

Die Vorgehensweise entspricht der von Fall 4 und wird daher nicht näher ausgeführt.

A N H A N G A2

I. Behauptung:

Wenn für die A-Matrix der Zustandsraumbeschreibung eine Diagonalmatrix $\underline{G}_1$ mit positiv reellen Elementen existiert, so daß

$$\underline{P}_1 = \underline{G}_1 - \underline{A}^T \underline{G}_1 \underline{A} \quad (\text{A2-1})$$

eine positiv definite Matrix ist, dann existiert im Fall der Zustandsraum-Struktur eine entsprechende Diagonalmatrix $\underline{G}$, so daß, bei geeigneter Wahl der Verknüpfungsmatrix $\underline{V}$ (siehe Gleichung (2-23)) mit

$$\underline{P} = \underline{G} - \underline{V}^T \underline{G} \underline{V} , \quad (\text{A2-2})$$

eine positiv semidefinite Matrix gegeben ist.

B_e_w_e_i_s_:

1. Bestimmung von $\underline{V}$:

Es wird zunächst der Zusammenhang zwischen der Zustandsraumbeschreibung und dem Realisierungsmodell nach Abschnitt 2.3.1 gesucht. Offenbar gilt für den ZV-Vektor:

$$\underline{a}_1(k) = \underline{x}(k) \quad (\text{A2-3a})$$

Die Eingangs- und Ausgangsfolgen erscheinen an unterschiedlichen Toren. Es ist deshalb sinnvoll, die Vektoren $\underline{a}_3(k)$ und $\underline{b}_3(k)$ nochmals in die Vektoren $\underline{a}_{31}(k)$, $\underline{a}_{32}(k)$ und $\underline{b}_{31}(k)$, $\underline{b}_{32}(k)$ zu unterteilen, so daß gilt:

$$\underline{a}_{31}(k) = \underline{v}(k) \quad \text{Eingangsfolgen-Vektor}$$

$$\underline{b}_{31}(k) = \underline{0} \quad (A2-3b)$$

$$\underline{a}_{32}(k) = \underline{0}$$

$$\underline{b}_{32}(k) = \underline{y}(k) \quad \text{Ausgangsfolgen-Vektor}$$

Berücksichtigt man schließlich noch, daß die Quantisierer unmittelbar vor den Verzögerungsgliedern angeordnet sind, lassen sich folgende Beziehungen angeben:

$$\underline{b}_2(k) = \underline{A} \underline{a}_1(k) + \underline{B} \underline{a}_{31}(k)$$

$$\underline{b}_1(k) = \underline{a}_2(k) \quad (A2-4)$$

$$\underline{b}_{31}(k) = \underline{0}$$

$$\underline{b}_{32}(k) = \underline{C} \underline{a}_1(k) + \underline{D} \underline{a}_{31}(k)$$

Die Matrix $\underline{V}$ besitzt demnach die Form

$$\underline{V} = \begin{bmatrix} \underline{0} & \underline{E} & \underline{0} & \underline{0} \\ \underline{A} & \underline{0} & \underline{B} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{C} & \underline{0} & \underline{D} & \underline{0} \end{bmatrix} \quad (A2-5)$$

wobei $\underline{E}$ eine Einheitsmatrix und $\underline{0}$ eine Null-Matrix mit jeweils passender Dimension ist.

2. Bestimmung von $\underline{P} = \underline{G} - \underline{V}^T \underline{G} \underline{V}$:

Zunächst unterteilen wir die Diagonalmatrix $\underline{G}$ gemäß

$$\underline{G} = \begin{bmatrix} \underline{G}_1 & & & \\ & \underline{G}_2 & & \\ & & \underline{G}_3 & \\ & & & \underline{G}_4 \end{bmatrix} \quad (A2-6)$$

in vier Unter-Diagonalmatrizen mit folgenden Dimensionen:

Dimension von $\underline{G}_1$ = Zahl der Komponenten von $\underline{a}_1, \underline{b}_1$

Dimension von $\underline{G}_2$ = Zahl der Komponenten von $\underline{a}_2, \underline{b}_2$

Dimension von $\underline{G}_3$ = Zahl der Komponenten von $\underline{a}_{31}, \underline{b}_{31}$

Dimension von $\underline{G}_4$ = Zahl der Komponenten von $\underline{a}_{32}, \underline{b}_{32}$

Damit errechnet man

$$\underline{P} = \underline{G} - \underline{V}^T \underline{G} \underline{V} = \begin{bmatrix} (\underline{G}_1 - \underline{A}^T \underline{G}_2 \underline{A} - \underline{C}^T \underline{G}_4 \underline{C}) & \underline{O} & (-\underline{A}^T \underline{G}_2 \underline{B} - \underline{C}^T \underline{G}_4 \underline{D}) & \underline{O} \\ \underline{O} & (\underline{G}_2 - \underline{G}_1) & \underline{O} & \underline{O} \\ (-\underline{B}^T \underline{G}_2 \underline{A} - \underline{D}^T \underline{G}_4 \underline{C}) & \underline{O} & (\underline{G}_3 - \underline{B}^T \underline{G}_2 \underline{B} - \underline{D}^T \underline{G}_4 \underline{D}) & \underline{O} \\ \underline{O} & \underline{O} & \underline{O} & \underline{G}_4 \end{bmatrix}$$

(A2-7)

Wir setzen nun $\underline{G}_1 = \underline{G}_2$. Dies ist zunächst eine völlig willkürliche Maßnahme, die aber in keinem Widerspruch zum Bisherigen steht, da bzgl. der gesamten Diagonalmatrix $\underline{G}$ noch keine Einschränkungen gemacht wurden. Damit geht Glg. (A2-7) über in

$$\underline{P} = \begin{bmatrix} \underline{P}_{11} & \underline{O} & \underline{P}_{13} & \underline{O} \\ \underline{O} & \underline{O} & \underline{O} & \underline{O} \\ \underline{P}_{13}^T & \underline{O} & \underline{P}_{33} & \underline{O} \\ \underline{O} & \underline{O} & \underline{O} & \underline{G}_4 \end{bmatrix}$$

mit

$$\underline{P}_{11} = \underline{G}_1 - \underline{A}^T \underline{G}_1 \underline{A} - \underline{C}^T \underline{G}_4 \underline{C} \quad (\text{A2-8})$$

$$\underline{P}_{13} = -\underline{A}^T \underline{G}_1 \underline{B} - \underline{C}^T \underline{G}_4 \underline{D}$$

$$\underline{P}_{33} = \underline{G}_3 - \underline{B}^T \underline{G}_1 \underline{B} - \underline{D}^T \underline{G}_4 \underline{D} .$$

3. Definitheit von $\underline{P}$:

Die Frage lautet, ob sich eine Diagonalmatrix $\underline{G}$ finden läßt, so daß $\underline{P}$ positiv semidefinit wird. Dazu betrachten wir zunächst die Untermatrix $\underline{P}_{11}$. Nach Voraussetzung (A2-1) gibt es ein $\underline{G}_1$, so daß $(\underline{G}_1 - \underline{A}^T \underline{G}_1 \underline{A})$ positiv definit ist.

Daher läßt sich durch die Wahl hinreichend kleiner positiver Elemente von $\underline{G}_4$ stets erreichen, daß auch $\underline{P}_{11}$ positiv definit bleibt. Ebenso ist es offenbar möglich, durch die Wahl genügend großer positiver Elemente von $\underline{G}_3$ die Matrix $\underline{P}_{33}$ positiv definit zu machen. Wir wollen $\underline{G}_3$ noch in eine Summe von positiven Diagonalmatrizen gemäß

$$\underline{G}_3 = \underline{G}_3^{(1)} + \underline{G}_3^{(2)} \quad (\text{A2-9})$$

zerlegen, wobei $\underline{G}_3^{(1)}$ und $\underline{G}_3^{(2)}$ die gleiche Dimension wie $\underline{G}_3$ besitzen. Damit gilt

$$\begin{aligned} \underline{P}_{33} &= \underline{G}_3^{(2)} + \underbrace{\underline{G}_3^{(1)} - \underline{B}^T \underline{G}_1 \underline{B} - \underline{D}^T \underline{G}_4 \underline{D}}_{:= \underline{P}_{33}^{(1)}} \quad (\text{A2-10}) \\ &:= \underline{P}_{33}^{(1)} \end{aligned}$$

$\underline{G}_3^{(1)}$ sei so gewählt, daß $\underline{P}_{33}^{(1)}$ positiv definit ist. Wir bilden nun die quadratische Form

$$\underline{a}^T \underline{P} \underline{a} \quad \text{mit} \quad \underline{a} = \begin{bmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \underline{a}_{31} \\ \underline{a}_{32} \end{bmatrix}.$$

Unter Berücksichtigung von (A2-8) und (A2-10) erhält man

$$\begin{aligned} \underline{a}^T \underline{P} \underline{a} &= \underline{a}_1^T \underline{P}_{11} \underline{a}_1 + 0 \cdot \underline{a}_2 + \underline{a}_{32}^T \underline{G}_4 \underline{a}_{32} + \underline{a}_{31}^T \underline{P}_{33}^{(1)} \underline{a}_{31} + \\ &+ 2 \underline{a}_1^T \underline{P}_{13} \underline{a}_{31} + \underline{a}_{31}^T \underline{G}_3^{(2)} \underline{a}_{31} \quad (\text{A2-11}) \end{aligned}$$

Da der Teilvektor $\underline{a}_2$ mit $\underline{0}$ gewichtet wird, kann $\underline{P}$ nicht positiv definit, sondern allenfalls positiv semidefinit sein. Wie gezeigt wurde, sind $\underline{P}_{33}^{(1)}$, $\underline{G}_4$ positiv definit oder können durch geeignete Wahl von $\underline{G}$ dazu gemacht werden. Somit folgt wegen $\underline{a}_{31}^T \underline{P}_{33}^{(1)} \underline{a}_{31} \geq 0$ und $\underline{a}_{32}^T \underline{G}_4 \underline{a}_{32} \geq 0$:

$$\underline{a}^T \underline{P} \underline{a} \geq \underline{a}_1^T \underline{P}_{11} \underline{a}_1 + 2 \underline{a}_1^T \underline{P}_{13} \underline{a}_{31} + \underline{a}_{31}^T \underline{G}_3^{(2)} \underline{a}_{31}. \quad (\text{A2-12})$$

$\underline{P}_{11}$ und $\underline{G}_3^{(2)}$ sind positiv definite Matrizen. Außerdem wurden noch keine Einschränkungen bzgl. der Diagonalmatrix $\underline{G}_3^{(2)}$ gemacht. Somit läßt sich durch die Wahl hinreichend großer positiver Diagonalelemente von $\underline{G}_3^{(2)}$ stets erreichen, daß die rechte Seite von (A2-12) für beliebige Vektoren $\underline{a}_1$ und $\underline{a}_{31}$ größer oder gleich Null wird.

Damit ist gezeigt, daß sich eine Diagonalmatrix $\underline{G}$ angeben läßt, derart daß $\underline{P} = \underline{G} - \underline{V}^T \underline{G} \underline{V}$ eine positiv semidefinite Matrix ist, was zu beweisen war.

II. Behauptung:

Wenn eine Diagonalmatrix $\underline{G}_1$ mit positiven Elementen existiert, so daß

$$\underline{P}_1 = \underline{G}_1 - \underline{A}^T \underline{G}_1 \underline{A}$$

eine positiv semidefinite Matrix ist, dann existiert eine entsprechende Diagonalmatrix $\underline{G}$ derart, daß

$$\underline{P}_* = \underline{G} - \underline{V}_*^T \underline{G} \underline{V}_*$$

positiv semidefinit ist.

$\underline{V}_*$ geht aus $\underline{V}$ hervor, wenn die Untermatrizen $\underline{C}$ und $\underline{D}$ in (A2-5) gleich Null gesetzt werden.

B e w e i s :

Der Beweis ergibt sich durch geringfügige Modifikation des Beweises in I. Diese bezieht sich lediglich auf die Matrix $\underline{P}_{11}$ in (A2-8). Da $\underline{P}_1 = \underline{G}_1 - \underline{A}^T \underline{G}_1 \underline{A}$ jetzt nur noch positiv semi-definit ist, könnte für $\underline{C} \neq \underline{O}$ nicht mehr sichergestellt werden, daß $\underline{P}_{11}$ positiv (semi-) definit ist. Da aber hier der Fall $\underline{C} = \underline{O}$ betrachtet wird, ist $\underline{P}_{11} = \underline{P}_1$ positiv semidefinit. In Beweis I konnte $\underline{P}_{11}$ zwar sogar positiv definit gemacht werden, was aber für die anschließenden Folgerungen ohne Belang ist.

A N H A N G A 3

THEORETISCHE GRUNDLAGEN ZUR BESCHREIBUNG DER AUSSTEUERUNG UND DES RUNDUNGSGRAUSCHENS

1. Aussteuerung und Skalierung

1.1 Aussteuerungskriterien

Im Fall der Zustandsraumstrukturen ist die Aussteuerung der Zustandsvariablen zu begrenzen. Dazu sind geeignete Kriterien zur Beurteilung des Dynamikbereiches erforderlich.

Wir wollen uns im folgenden auf Systeme mit einem Eingang und einem Ausgang beschränken. Aus der Zustandsraumbeschreibung

$$\underline{x}(k) = \underline{A} \underline{x}(k) + \underline{b} v(k) \quad (\text{A3-1})$$

gewinnt man zunächst die Impulsantworten $f_v(k)$, $v = 1(1)n$, vom Filtereingang zu den Zustandsvariablen x_v als

$$\underline{f}(k) = \begin{bmatrix} f_1(k) \\ \vdots \\ f_n(k) \end{bmatrix} = \begin{matrix} 0 & \text{für } k \leq 0 \\ \underline{A}^{k-1} \underline{b} & \text{für } k \geq 1 \end{matrix} \quad (\text{A3-2})$$

Bei beliebiger Erregung $v(k)$ gilt dann

$$x_v(k) = \sum_{\kappa=0}^{\infty} f_v(\kappa) v(k-\kappa) \quad (\text{A3-3})$$

Mit einer auf $|v(k)| \leq 1$ beschränkten Eingangsfolge folgt die Abschätzung /6/

$$|x_v(k)| \leq \sum_{\kappa=0}^{\infty} |f_v(\kappa)| \quad (\text{A3-4})$$

Um die maximale Amplitude der Zustandsvariablen mit Sicherheit auf 1 zu beschränken, muß die in (A3-4) angegebene Norm durch geeignete Maßnahmen gleich 1 gemacht werden. In praktischen Fällen

ist die in (A3-4) bestimmte obere Schranke jedoch meist zu pessimistisch gewählt, so daß eine Skalierung nach diesem Kriterium zu einer schlechten Ausnutzung des verfügbaren Dynamikbereiches und damit zu einer Erhöhung des Rundungsrauschens führt. Daher wurde von Jackson /11/ zur Beurteilung der Aussteuerung die L_p -Norm herangezogen, die für eine Funktion $A(e^{j\Omega})$ wie folgt definiert ist:

$$\|A\|_p = \left[\frac{1}{2\pi} \int_0^{2\pi} |A(e^{j\Omega})|^p d\Omega \right]^{1/p}, \quad p \in \mathbb{N}. \quad (\text{A3-5})$$

Ist nun $V(z)$ die Z -Transformierte der Eingangsfolge $v(k)$ und $F_v(z)$ die Übertragungsfunktion vom Eingang zur v -ten Zustandsvariablen, so gilt gemäß der Hölderschen Ungleichung

$$|x_v(k)| \leq \|F_v\|_p \cdot \|v\|_q \quad \forall k \quad (\text{A3-6})$$

mit $\frac{1}{p} + \frac{1}{q} = 1 \quad \text{und} \quad p, q \geq 1.$

Unter der Annahme eines Eingangssignals, das die Bedingung $\|v\|_q = 1$ erfüllt, ergibt sich nach (A3-6) die obere Schranke

$$|x_v(k)| \leq \|F_v\|_p, \quad \frac{1}{p} + \frac{1}{q} = 1. \quad (\text{A3-7})$$

Von besonderer praktischer Bedeutung sind die Kombinationen $(p, q) = (\infty, 1)$ und $(p, q) = (2, 2)$. Für den Grenzfall der L_∞ -Norm ist nach /11/

$$\|F_v\|_\infty = \max_{\Omega} \{|F_v(e^{j\Omega})|\} \quad (\text{A3-8})$$

Wird durch geeignete Skalierung des Filters sichergestellt, daß $\|F_v\|_\infty = 1$ gilt, so ist entsprechend der Definition des Frequenzganges bei einer für alle Werte von k andauernden sinusförmigen Erregung mit der Amplitude $\hat{v}$ die maximale Amplitude der Zustandsvariablen $x_v(k)$ ebenfalls auf den Wert $\hat{v}$ beschränkt.

Im übrigen nimmt die L_p -Norm für $p = \infty$ ihren größten Wert an und stellt somit die strengste Skalierungsbedingung dar.

Die Kombination $(p,q) = (2,2)$ ist insbesondere wegen ihrer besseren analytischen Handhabbarkeit von Interesse. Gemäß Definition (A3-5) und unter Berücksichtigung der Parseval-Beziehung gilt

$$||F_v||_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |F_v(e^{j\Omega})|^2 d\Omega = \sum_{k=0}^{\infty} f_v^2(k) \quad (A3-9)$$

Zur Interpretation dieser Norm sei am Filtereingang ein mittelwertfreies weißes Geräusch der Varianz σ_v^2 angenommen. Dann sind die Zustandsvariablen ebenfalls rauschartig und besitzen die Varianz $\sigma_{x_v}^2 = ||F_v||_2^2 \cdot \sigma_v^2$. Bei Gauß-Verteilung läßt sich hieraus die Wahrscheinlichkeit für das Auftreten von Überläufen berechnen.

Als Skalierungsbedingung wird

$$\delta \cdot ||F_v||_2^2 = 1 \quad (A3-10)$$

gefordert, wobei mit der Konstanten δ ein zusätzlicher "Sicherheitsfaktor" vorgebar ist.

Im Zusammenhang mit der Minimierung des Rundungsrauschens ist die Kovarianzmatrix $\underline{K}$ der Impulsantworten $f_v(k)$ von Bedeutung /19,21/. Die Elemente dieser Matrix sind von der Form

$$K_{\mu v} = \sum_{k=0}^{\infty} f_{\mu}(k) f_v(k) \quad (A3-11)$$

In der Hauptdiagonalen erscheinen die quadrierten L_2 -Normen der Zustandsvariablen. Somit gilt u.a.

$$\text{Spur } \{\underline{K}\} = \sum_{v=1}^n ||F_v||_2^2 \quad (A3-12)$$

Unter Beachtung von (A3-2) ergibt sich $\underline{K}$ als

$$\begin{aligned} \underline{K} &= \sum_{k=1}^{\infty} \underline{f}(k) \underline{f}^T(k) = \sum_{k=1}^{\infty} \begin{bmatrix} \underline{A}^{k-1} & \underline{b} \end{bmatrix} \begin{bmatrix} \underline{A}^{k-1} & \underline{b} \end{bmatrix}^T \\ &= \sum_{k=0}^{\infty} \begin{bmatrix} \underline{A}^k & \underline{b} \end{bmatrix} \begin{bmatrix} \underline{A}^k & \underline{b} \end{bmatrix}^T \end{aligned} \quad (A3-13)$$

Ein geeignetes Verfahren zur numerischen Berechnung von $\underline{K}$ ist in /21/ angegeben.

Offenbar ist $\underline{K}$ bei gegebener Filterübertragungsfunktion nicht eindeutig festgelegt, da $\underline{A}$ und $\underline{b}$ sich nicht eindeutig ergeben. Wie in Abschnitt 2.2.1 erläutert, läßt sich der ZV-Vektor $\underline{x}$ mit Hilfe einer nichtsingulären ($n \times n$) Matrix $\underline{T}$ gemäß $\underline{x}' = \underline{T}^{-1} \underline{x}$ transformieren. Durch Einsetzen der geänderten Matrizen $\underline{A}' = \underline{T}^{-1} \underline{A} \underline{T}$ bzw. $\underline{A}'^k = \underline{T}^{-1} \underline{A}^k \underline{T}$ und $\underline{b}' = \underline{T}^{-1} \underline{b}$ bestimmt man aus (A3-13) die neue Kovarianzmatrix

$$\underline{K}' = \underline{T}^{-1} \underline{K} \underline{T}^{-T} \quad (\text{A3-14})$$

1.2 Skalierung

Es werden zwei Möglichkeiten angegeben, die im Zusammenhang mit den betrachteten Filterstrukturen von Bedeutung sind:

a) Transformation mit einer Diagonalmatrix

Die Zustandsraum-Beschreibungen des unskalierten bzw. des skalierten Systems seien:

$$\begin{aligned} \text{unskaliert:} \quad \underline{x}(k+1) &= \underline{A} \underline{x}(k) + \underline{b} v(k) \\ y(k) &= \underline{c}^T \underline{x}(k) + d v(k) \end{aligned} \quad (\text{A3-15a})$$

$$\begin{aligned} \text{skaliert:} \quad \underline{x}'(k+1) &= \underline{A}' \underline{x}'(k) + \underline{b}' v(k) \\ y(k) &= \underline{c}'^T \underline{x}'(k) + d v(k) \end{aligned} \quad (\text{A3-15b})$$

Wir fordern z.B. eine Skalierung nach der L_∞ -Norm, so daß für das skalierte Filter

$$\|F_v'\|_\infty = 1, \quad v = 1(1)n \quad (\text{A3-16})$$

gilt. Ist nun $||F_v||_\infty$ die L_∞ -Norm der unskalierten Zustandsvariablen x_v , so ist

$$\underline{x}' = \underbrace{\text{Diag} \{ ||F_1||_\infty^{-1}, \dots, ||F_n||_\infty^{-1} \}}_{:= \underline{T}^{-1}} \underline{x} \quad (\text{A3-17})$$

zu wählen. Die Transformation mit der so definierten Diagonalmatrix $\underline{T}$ liefert

$$\underline{A}' = \underline{T}^{-1} \underline{A} \underline{T}, \quad \underline{b}' = \underline{T}^{-1} \underline{b}, \quad \underline{c}'^T = \underline{c}^T \underline{T}. \quad (\text{A3-18})$$

Bei Verwendung anderer Normen ist die Vorgehensweise entsprechend.

b) Transformation mit einer Orthogonalmatrix

Wir beschränken uns im folgenden auf eine Skalierung nach der L_2 -Norm, wobei

$$||F_v'||_2 = 1/\delta \quad (\text{A3-19})$$

gelten soll. D.h. die Kovarianzmatrix $\underline{K}'$ des skalierten Filters besitzt die Hauptdiagonalelemente

$$K'_{vv} = \frac{1}{\delta^2}, \quad v = 1(1)n \quad (\text{A3-20})$$

und es gilt

$$\text{Spur} \{ \underline{K}' \} = n \frac{1}{\delta^2} \quad (\text{A3-21})$$

Wir legen nun ein unskaliertes System zugrunde, dessen $\underline{K}$ -Matrix bereits die Eigenschaft

$$\text{Spur} \{ \underline{K} \} = n \frac{1}{\delta^2} \quad (\text{A3-22})$$

besitzt. Dies läßt sich vorab leicht durch Transformation mit einer geeigneten skalaren Konstanten erreichen. Die Hauptdiagonal-

elemente $K_{\nu\nu}$ sind i.a. aber noch nicht gleich.

Bei Transformation mit einer orthogonalen Matrix $\underline{T} = \underline{U}$ folgt gemäß (A3-14)

$$\underline{K}' = \underline{U}^{-1} \underline{K} \underline{U} = \underline{U}^T \underline{K} \underline{U} \quad (\text{A3-23})$$

$\underline{U}$ ist nun derart zu bestimmen, daß die Skalierungsbedingung (A3-20) erfüllt wird.

Nach einem Vorschlag von Mullis und Robert /21/ spaltet man $\underline{U}$ in ein Produkt von Matrizen $\underline{U}_i$ auf, welche nur jeweils eine zwei-dimensionale Drehung bewirken, z.B.

$$\underline{U}_i = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & 1 & & & \vdots \\ \vdots & & & \cos\theta & \sin\theta & \vdots \\ \vdots & & & & 1 & \vdots \\ \vdots & & & & & \ddots \\ 0 & \dots & \dots & \dots & \dots & 0 & 1 \end{bmatrix} \quad (\text{A3-24})$$

μ ν

Unterwirft man die Matrix $\underline{K}$ einer Transformation mit $\underline{U}_i$, so werden in der Hauptdiagonalen nur die Elemente $K_{\mu\mu}$ und $K_{\nu\nu}$ verändert. Durch Wahl eines geeigneten Drehwinkels θ lassen sich diese Hauptdiagonalelemente jeweils gleich machen, wie an einem Beispiel zweiten Grades erläutert wird:

$$\underline{K}' = \underline{U}^T \underline{K} \underline{U} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{12} & K_{22} \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Für

$$\tan 2\theta = \frac{K_{11} - K_{22}}{2K_{12}} \quad (\text{A3-25})$$

wird

$$K'_{11} = K'_{22} = \frac{1}{2} [K_{11} + K_{22}] .$$

In gleicher Weise kann mit höherdimensionalen Matrizen $\underline{K}$ verfahren werden, wobei jeweils die größten und kleinsten Diagonalelemente durch ihr arithmetisches Mittel ersetzt werden. Bei fortgesetzter Anwendung dieser Prozedur konvergieren nach /21/ alle Diagonalelemente gegen den gleichen Wert. Die gesamte Orthogonalmatrix $\underline{U}$ ergibt sich zu

$$\underline{U} = \dots \underline{U}_4 \underline{U}_3 \underline{U}_2 \underline{U}_1 . \quad (A3-26)$$

Da die Ähnlichkeitstransformation $\underline{K}' = \underline{U}^T \underline{K} \underline{U}$ die Spur von $\underline{K}$ nicht beeinflußt /20/ folgt mit (A3-21)

$$K'_{\mu\mu} = K'_{\nu\nu} = \frac{1}{n} \text{Spur} \{K\} = \frac{1}{\delta 2}$$

$$\text{bzw.} \quad ||F'_v||_2 = 1/\delta, \quad v = 1(1)n .$$

Die Systemmatrizen des skalierten Filters ergeben sich zu

$$\underline{A}' = \underline{U}^T \underline{A} \underline{U}, \quad \underline{b}' = \underline{U}^T \underline{b}, \quad \underline{c}'^T = \underline{c}^T \underline{U}. \quad (A3-27)$$

2. Analyse des Rundungsrauschens

Infolge der nach Multiplikationen erforderlichen Wortlängenverkürzung entstehen Quantisierungsfehler. Dies führt bei nichtperiodischen Eingangssignalen zu einem rauschartigen Störanteil am Filterausgang. Theoretische Modelle zur Analyse dieses Rundungsrauschens wurden in der Vergangenheit ausführlich diskutiert. Eine zusammenfassende Darstellung findet sich z.B. in /5/, wobei unterschiedliche Methoden der Wortlängenverkürzung zugrundegelegt werden. Im Rahmen der vorliegenden Arbeit wird nur der Fall der normalen Zweierkomplement-Rundung betrachtet. Die Untersuchungen werden mit Hilfe eines statistischen Modells durchgeführt, dessen Brauchbarkeit durch Simulation und Messung an realisierten Festkomma-Filtern vielfach bestätigt werden konnte. Dabei ersetzt man

die Quantisierer durch Rauschquellen, die ein mittelwertfreies in $[-\frac{Q}{2}, \frac{Q}{2}]_2$ gleichverteiltes und weißes Geräusch der Varianz $\sigma_o^2 = \frac{Q^2}{12}$ (Q = Quantisierungsintervall) liefern. Bild A3.1 erläutert dieses Modell für ein Zustandsraumfilter 2. Grades, wobei wie in Bild 2.2 die Rundung nach Addition der Teilprodukte erfolgt.

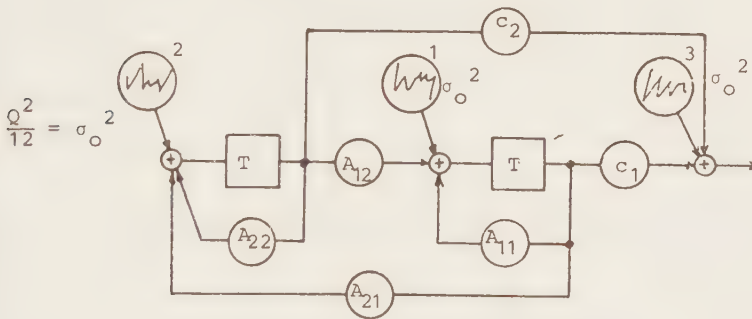


Bild A3.1: Statistisches Modell zur Beschreibung des Rundungsrauschens

Ferner wird angenommen, daß die einzelnen Rauschquellen nicht miteinander korreliert sind. Die Berechnung der Rauschleistung am Filterausgang darf daher für die einzelnen Störquellen getrennt erfolgen. Die Störleistungsanteile addieren sich.

Bezeichnet $g_v(k)$ die Impulsantwort von der v -ten Störquelle zum Filterausgang, dann bestimmt man die von dieser Rauschquelle verursachte Fehlerleistung σ_v^2 zu

$$\sigma_v^2 = \sigma_o^2 \sum_{k=0}^{\infty} g_v^2(k) . \quad (A3-28)$$

Die unmittelbar am Filterausgang auftretende Störquelle (Störquelle Nr. 3 in Bild A3.1) nimmt eine Sonderstellung ein. Ihr Beitrag zum Rundungsrauschen ist wegen der fehlenden Rückkopplung im allgemeinen gering. Er kann außerdem durch Transformation der

Struktur nicht beeinflußt werden. Diese Rundungsstelle bleibt im folgenden zunächst unberücksichtigt.

Wir betrachten nun die für den Entwurf rauschoptimaler Filter wichtige Kovarianzmatrix $\underline{W}$, deren Elemente durch

$$\underline{W}_{\mu\nu} = \sum_{k=0}^{\infty} g_{\mu}(k) g_{\nu}(k) \quad (\text{A3-29})$$

definiert sind. Faßt man die $g_{\nu}(k)$ in einem Zeilenvektor $\underline{g}(k)$ zusammen, dann folgt aus der Zustandsraumbeschreibung (Abschnitt 2.2)

$$\underline{g}(k) = \begin{array}{ll} \underline{c}^T \underline{A}^{k-1} & ; \quad k \geq 1 \\ 0 & ; \quad k \leq 0 \end{array} \quad (\text{A3-30})$$

Unter Berücksichtigung der Definition (A3-29) erhält man

$$\underline{W} = \sum_{k=0}^{\infty} (\underline{c}^T \underline{A}^k)^T \underline{c}^T \underline{A}^k . \quad (\text{A3-31})$$

Gemäß (A3-28) und (A3-29) ergibt sich die Störleistung am Filterausgang aus den Hauptdiagonalelementen von $\underline{W}$.

Unter der Annahme, daß die Rundung jeweils nach Zusammenfassen der Teilprodukte erfolgt, erhält man für die Varianz des Rauschens

$$\sigma'^2 = \sum_{\nu} \sigma_{\nu}^2 = \sigma_0^2 \text{ Spur } \{\underline{W}\} . \quad (\text{A3-32})$$

Berücksichtigt man zusätzlich die Rundungsstelle am Filterausgang, so geht (A3-32) über in die Form

$$\sigma^2 = \sigma'^2 + \sigma_0^2 = \sigma_0^2 [\text{Spur } \{\underline{W}\} + 1] . \quad (\text{A3-33})$$

Es sei noch erwähnt, daß eine Transformation des ZV-Vektors gemäß $\underline{x}' = \underline{T}^{-1} \underline{x}$ zu einer modifizierten Matrix

$$\underline{W}' = \underline{T}^T \underline{W} \underline{T} \quad (\text{A3-34})$$

führt. Die einfache Herleitung erfolgt wie im Fall der $\underline{K}$ -Matrix (siehe Glg. (A3-14)).

A N H A N G A 4

ENTWURF DER NORMALFORM-STRUKTUR BEI GEGEBENER ÜBERTRAGUNGS-FUNKTION

1. Entwurf der allgemeinen Struktur

Wir betrachten ein Filter mit einem Eingang und einem Ausgang. Die zugehörige Zustandsraum-Darstellung lautet (Abschnitt 2.2.1):

$$\underline{x}(k+1) = \underline{A} \underline{x}(k) + \underline{b} v(k) \quad (A4-1)$$

$$y(k) = \underline{c}^T \underline{x}(k) + d v(k)$$

$\underline{b}$: $(n \times 1)$ Spaltenvektor

$\underline{c}^T$: $(1 \times n)$ Zeilenvektor

Nach Abschnitt 3.2.2, Glg. (3-2) besitzt eine normale Matrix $\underline{A}$ einen orthogonalen Satz von Eigenvektoren $\underline{\phi}_\nu$, $\nu = 1(1)n$, mit

$$\begin{aligned} \underline{\phi}^{*T} \underline{\phi}_\nu &= \delta_{\mu\nu} = \begin{cases} 1 & , \mu = \nu \\ 0 & , \mu \neq \nu \end{cases} \end{aligned} \quad (A4-2)$$

Damit ergibt sich die Spektralzerlegung

$$\underline{A} = \sum_{\nu=1}^n \lambda_\nu \underline{\phi}_\nu \underline{\phi}_\nu^{*T} . \quad (A4-3)$$

Man überprüft die Gültigkeit dieser Beziehung leicht durch rechtsseitige Multiplikation mit einem beliebigen Eigenvektor $\underline{\phi}_\mu$, $\mu \in \{1, 2, \dots, n\}$. Da die $\underline{\phi}_\mu$ orthonormal sind, geht (A4-3) dann über in

$$\underline{A} \underline{\phi}_\mu = \lambda_\mu \underline{\phi}_\mu \quad \mu = 1(1)n .$$

Die $\underline{\phi}_\nu$ bilden eine Basis im $\mathbb{R}^n$. Somit lassen sich ferner die Vektoren $\underline{b}$ und $\underline{c}^T$ der Zustandsraum-Beschreibung durch die Linearkombinationen

$$\begin{aligned}\underline{b} &= \sum_{v=1}^n \beta_v \underline{\phi}_v \\ \underline{c}^T &= \sum_{v=1}^n \gamma_v \underline{\phi}_v^{*T}\end{aligned}\tag{A4-4}$$

darstellen, wobei die Koeffizienten β_v, γ_v i.a. komplex sind.
Wir setzen nun (A3-3) und (A3-4) in die Übertragungsfunktion

$$H(z) = d + \underline{c}^T [z\underline{E} - \underline{A}]^{-1} \underline{b}\tag{A4-5}$$

ein.

Dazu drücken wir zunächst die Einheitsmatrix $\underline{E}$ durch die ortho-normalen Eigenvektoren von $\underline{A}$ aus:

$$\underline{E} = \sum_{v=1}^n \underline{\phi}_v \cdot \underline{\phi}_v^{*T}\tag{A4-6}$$

Mit (A4-3) folgt

$$z\underline{E} - \underline{A} = \sum_{v=1}^n [z - \lambda_v] \underline{\phi}_v \underline{\phi}_v^{*T}$$

und

$$[z\underline{E} - \underline{A}]^{-1} = \sum_{v=1}^n \frac{1}{z - \lambda_v} \underline{\phi}_v \underline{\phi}_v^{*T}.\tag{A4-7}$$

Zur Kontrolle bilden wir

$$\begin{aligned}[z\underline{E} - \underline{A}]^{-1} \cdot [z\underline{E} - \underline{A}] &= \sum_{\mu=1}^n \sum_{v=1}^n \frac{z - \lambda_v}{z - \lambda_\mu} \underline{\phi}_\mu \underline{\phi}_\mu^{*T} \underline{\phi}_v \underline{\phi}_v^{*T} \\ &= \sum_{v=1}^n \underline{\phi}_v \underline{\phi}_v^{*T} = \underline{E}.\end{aligned}$$

Durch Einsetzen von (A4-4) und (A4-7) in (A4-5) gewinnt man schließlich

$$\begin{aligned}H(z) &= d + \sum_{\mu=1}^n \sum_{v=1}^n \sum_{\kappa=1}^n \frac{\gamma_\mu \beta_\kappa}{z - \lambda_v} \underline{\phi}_\mu^{*T} \underline{\phi}_v \underline{\phi}_v^{*T} \underline{\phi}_\kappa \\ &= d + \sum_{v=1}^n \frac{R_v}{z - \lambda_v}; \quad R_v := \gamma_v \beta_v\end{aligned}\tag{A4-8}$$

$H(z)$ ist unabhängig von den Eigenvektoren $\underline{\phi}_v$. D.h. jeder beliebige Satz von n orthonormalen Vektoren $\underline{\phi}_v$ liefert mit (A4-3), (A4-4) geeignete Systemmatrizen, welche die gewünschte Übertragungsfunktion realisieren. Allerdings ist sicherzustellen, daß $\underline{A}$, $\underline{b}$ und $\underline{c}^T$ reell sind. Dazu wählt man die Eigenvektoren $\underline{\phi}_v$ in Abhängigkeit von den Eigenwerten λ_v entweder reell oder als konjugiert komplexe Paare. Ferner müssen die β_v und γ_v , passend zu den Eigenvektoren, reell oder konjugiert komplex sein.

Diese Forderungen sind offenbar stets erfüllbar. Die darüber hinaus verbleibenden Freiheitsgrade werden für die Skalierung und die Minimierung des Rundungsrauschens genutzt, wie im folgenden gezeigt wird.

Ziel ist es nun, für den Fall L_2 -skalierter Zustandsvariabler das Rundungsrauschen zu minimieren. Dazu betrachten wir zunächst die Kovarianzmatrix $\underline{K}$ der Impulsantworten von Filtereingang zu den Zustandsgrößen. Nach Anhang A3, Glg. (A3-13) gilt

$$\underline{K} = \sum_{k=0}^{\infty} (\underline{A}^k \underline{b}) (\underline{A}^k \underline{b})^T = \sum_{k=0}^{\infty} \underline{A}^k \underline{b} \underline{b}^T \underline{A}^{kT}. \quad (\text{A4-9})$$

Mit einer $\underline{A}$ -Matrix der Form (A4-3) und bei Beachtung der Orthonormalität der $\underline{\phi}_v$ erhält man

$$\underline{A}^k = \sum_{v=1}^n \lambda_v^k \underline{\phi}_v \underline{\phi}_v^{*T} \quad (\text{A4-10})$$

bzw.

$$\underline{A}^{kT} = (\underline{A}^k)^{*T} = \sum_{v=1}^n \lambda_v^{*k} \underline{\phi}_v \underline{\phi}_v^{*T}.$$

Entsprechend gilt nach (A4-4)

$$\underline{b} = \sum_{v=1}^n \beta_v \underline{\phi}_v, \quad \underline{b}^T = \underline{b}^{*T} = \sum_{v=1}^n \beta_v^* \underline{\phi}_v^{*T}. \quad (\text{A4-11})$$

Durch Einsetzen von (A4-10), (A4-11) in (A4-9) ergibt sich

$$\begin{aligned}
 \underline{K} &= \sum_{k=0}^{\infty} \sum_{\mu=1}^n \sum_{\nu=1}^n \sum_{\kappa=1}^n \sum_{\xi=1}^n \lambda_{\mu}^k \beta_{\nu}^* \lambda_{\xi}^{*k} \underbrace{\phi_{\mu}^{\ast T} \phi_{\nu}}_{\delta_{\mu\nu}} \underbrace{\phi_{\kappa}^{\ast T} \phi_{\xi}}_{\delta_{\kappa\xi}} \phi_{\xi}^{\ast T} \\
 &= \sum_{k=0}^{\infty} \sum_{\mu=1}^n \sum_{\kappa=1}^n \lambda_{\mu}^k \lambda_{\kappa}^{*k} \beta_{\mu}^* \beta_{\kappa}^* \phi_{\mu}^{\ast T} \phi_{\kappa}^{\ast T} \\
 &= \sum_{\mu=1}^n \sum_{\kappa=1}^n \left[\sum_{k=0}^{\infty} (\lambda_{\mu} \lambda_{\kappa}^*)^k \right] \cdot \beta_{\mu}^* \beta_{\kappa}^* \phi_{\mu}^{\ast T} \phi_{\kappa}^{\ast T} \\
 &= \sum_{\mu=1}^n \sum_{\kappa=1}^n \frac{\beta_{\mu}^* \beta_{\kappa}^*}{1 - \lambda_{\mu} \lambda_{\kappa}^*} \phi_{\mu}^{\ast T} \phi_{\kappa}^{\ast T} .
 \end{aligned} \tag{A4-12}$$

Von Interesse ist noch die Spur von $\underline{K}$. Sie ist gleich der Summe der Quadrate der L_2 -Normen aller Zustandsvariablen:

$$\text{Spur}\{\underline{K}\} = \sum_{\mu=1}^n \sum_{\nu=1}^n \frac{\beta_{\mu}^* \beta_{\nu}^*}{1 - \lambda_{\mu} \lambda_{\nu}^*} \text{Spur} \{ \phi_{\mu}^{\ast T} \phi_{\nu}^{\ast T} \} .$$

Wegen (/20/)

$$\text{Spur} \{ \phi_{\mu}^{\ast T} \phi_{\nu}^{\ast T} \} = \text{Spur} \{ \phi_{\nu}^{\ast T} \phi_{\mu}^{\ast T} \} = \delta_{\mu\nu}$$

folgt

$$\text{Spur} \{ \underline{K} \} = \sum_{\nu=1}^n \frac{|\beta_{\nu}|^2}{1 - |\lambda_{\nu}|^2} . \tag{A4-13}$$

Aus der Skalierungsvorschrift $K_{\nu\nu} = 1/\delta^2$, $\nu = 1(1)n$, resultiert die notwendige Bedingung

$$\text{Spur} \{ \underline{K} \} = n \frac{1}{\delta^2}$$

bzw.

$$\sum_{\nu=1}^n \frac{|\beta_{\nu}|^2}{1 - |\lambda_{\nu}|^2} = n \frac{1}{\delta^2} . \tag{A4-14}$$

Im nächsten Schritt soll ein Ausdruck für die Leistung des Rundungsrauschens ermittelt werden. Dazu berechnen wir die in Anhang A3, Glg. (A3-31) angegebene Kovarianzmatrix

$$\underline{W} = \sum_{k=0}^{\infty} (\underline{C}^T \underline{A}^k)^T \cdot (\underline{C}^T \underline{A}^k) .$$

Durch Einsetzen von (A4-3) und (A4-4) gewinnt man die (A4-12) entsprechende Beziehung

$$\underline{W} = \sum_{\mu=1}^n \sum_{v=1}^n \frac{\gamma_{\mu}^* \gamma_v}{1 - \lambda_{\mu}^* \lambda_v} \phi_{\mu} \phi_v^* \quad (\text{A4-15})$$

Mit Anhang A3, Glg. (A3-32), ergibt sich die Rauschleistung σ'^2 aus der Spur von $\underline{W}$ als

$$\sigma'^2 = \sigma_0^2 \text{ Spur } \{\underline{W}\} = \sigma_0^2 \sum_{v=1}^n \frac{|\gamma_v|^2}{1 - |\lambda_v|^2} . \quad (\text{A4-16})$$

σ'^2 soll nun unter der Nebenbedingung (A4-14) minimiert werden, wobei nach (A4-8) die Residuen $R_v = \beta_v \gamma_v$, $v = 1(1)n$, durch die Partialbruchentwicklung von $H(z)$ festgelegt sind.

Wir lösen das Problem mit dem Verfahren der Lagrangeschen Multiplikatoren /37/ und bilden zunächst die Summe

$$\frac{\sigma'^2}{\sigma_0^2} + \xi^2 \frac{n}{\delta^2} ,$$

wobei ξ eine vorerst beliebige positive Konstante ist. Mit (A4-14), (A4-16) und $\beta_v = R_v/\gamma_v$ erhält man

$$\frac{\sigma'^2}{\sigma_0^2} + \xi^2 \frac{n}{\delta^2} = \sum_{v=1}^n \frac{|\gamma_v|^2 + \xi^2 \frac{|R_v|^2}{|\gamma_v|^2}}{1 - |\lambda_v|^2} . \quad (\text{A4-17})$$

σ'^2 wird offenbar minimal, wenn der Summenausdruck (A4-17) minimal wird. Die partielle Ableitung nach der Variablen $|\gamma_{\mu}|^2$ liefert

$$\frac{\partial}{\partial |\gamma_{\mu}|^2} \left(\frac{\sigma'^2}{\sigma_0^2} + \xi^2 \frac{n}{\delta^2} \right) = \frac{1 - \frac{\xi^2 |R_{\mu}|^2}{|\gamma_{\mu}|^4}}{1 - |\lambda_{\mu}|^2} ; \mu = 1(1)n .$$

Daraus erhält man die Extremalstellen

$$|\gamma_\mu|^2 = \xi |R_\mu| \quad ; \quad \mu = 1(1)n . \quad (A4-18)$$

Die Konstante ξ ermittelt man aus der Nebenbedingung (A4-14) mit $|\beta_v|^2 = |R_v|^2 / |\gamma_v|^2 = |R_v| / \xi$:

$$\sum_{v=1}^n \frac{|R_v| / \xi}{1 - |\lambda_v|^2} = n \frac{1}{\delta^2} \quad \xi = \frac{\delta^2}{n} \sum_{v=1}^n \frac{|R_v|}{1 - |\lambda_v|^2} .$$

Zusammen mit (A4-18) ergibt sich das gewünschte Resultat

$$|\gamma_\mu|^2 = \frac{\delta^2}{n} |R_\mu| \cdot \sum_{v=1}^n \frac{|R_v|}{1 - |\lambda_v|^2} \quad ; \quad \mu = 1(1)n \quad (A4-19)$$

$$\beta_\mu = R_\mu / \gamma_\mu$$

Durch Einsetzen in (A4-16) erhält man die minimale Geräuschleistung

$$\sigma_{\min}^2 = \sigma_o^2 \frac{\delta^2}{n} \left[\sum_{v=1}^n \frac{|R_v|}{1 - |\lambda_v|^2} \right]^2 \quad (A4-20)$$

Die Wahl der γ_μ und β_μ gemäß (A4-17) gewährleistet i.a. noch nicht, daß alle Zustandsvariablen gleichmäßig angesteuert sind. Der Grund ist, daß die Skalierungsbedingung (A4-14) notwendig aber nicht hinreichend ist. Abhilfe schafft hier die Transformation des ZV-Vektors $\underline{x}$ mit einer geeigneten orthogonalen Matrix $\underline{U}$ (siehe Anhang A3, Abschnitt 1.2). Damit ergibt sich eine neue $\underline{A}$ -Matrix $\underline{A}' = \underline{U}^T \underline{A} \underline{U}$. $\underline{A}$ war normal und erfüllt daher die notwendige und hinreichende Bedingung $\underline{A}^T \underline{A} = \underline{A} \underline{A}^T$. Wegen $\underline{A}'^T \underline{A}' = \underline{U}^T \underline{A}^T \underline{U} \underline{U}^T \underline{A} \underline{U} = \underline{U}^T \underline{A}^T \underline{A} \underline{U} = \underline{U}^T \underline{A} \underline{A}^T \underline{U} = \underline{A}' \underline{A}'^T$ ist $\underline{A}'$ ebenfalls normal. Die Kovarianzmatrix $\underline{w}$ transformiert sich gemäß Anhang A3, Glg. (A3-34), in

$$\underline{w}' = \underline{U}^T \underline{w} \underline{U} = \underline{U}^{-1} \underline{w} \underline{U} . \quad (A4-21)$$

Da eine derartige Ähnlichkeitstransformation die Spur von $\underline{w}$ nicht verändert /20/, bleibt die minimale Geräuschleistung ebenfalls erhalten.

Für die praktische Anwendung werden noch einmal die Entwurfs-schritte zusammengestellt:

Wir setzen ein Filter mit einem Eingang und einem Ausgang voraus, dessen Übertragungsfunktion nur einfache Polstellen besitzt.

Schritt 1:

Man bestimmt die Partialbruchentwicklung der Übertragungsfunktion $H(z)$. Damit sind nach Glg. (A4-8) die Produkte $\beta_\nu \gamma_\nu = R_\nu$ festgelegt.

Schritt 2:

Es sollen nun als Zwischenergebnis eine Matrix $\underline{A}$ sowie Vektoren $\underline{b}$ und $\underline{c}^T$ gemäß (A4-3) und (A4-4) angegeben werden.

Zunächst bestimmt man aus (A4-19) die Beträge der Koeffizienten γ_μ . Wir wählen nun einen Satz von n orthonormalen Eigenvektoren $\underline{\phi}_\mu$ derart, daß reellen Eigenwerten λ_μ reelle $\underline{\phi}_\mu$ und konjugiert komplexen Eigenwerten konjugiert komplexe Paare von Eigenvektoren zugeordnet werden können. Im übrigen ist die Wahl der $\underline{\phi}_\mu$ beliebig. Ebenso verfahren wir mit den Koeffizienten γ_μ , von denen lediglich ihr Betrag vorgeschrieben ist. Die Koeffizienten β_μ gewinnt man gemäß $\beta_\mu = R_\mu / \gamma_\mu$. Damit ist $\underline{A}$, $\underline{b}$ und $\underline{c}^T$ angebbar. Ferner läßt sich die Kovarianzmatrix $\underline{K}$ nach (A4-12) ermitteln.

Als Beispiel werden für ein Filter 6-ten Grades mit 4 komplexen und 2 reellen Eigenwerten geeignete Eigenvektoren angegeben:

$$\underline{\phi}_1 = \sqrt{\frac{1}{2}} \begin{bmatrix} 1 \\ j \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \underline{\phi}_2 = \underline{\phi}_1^*, \quad \underline{\phi}_3 = \sqrt{\frac{1}{2}} \begin{bmatrix} 0 \\ 0 \\ 1 \\ j \\ 0 \\ 0 \end{bmatrix}, \quad \underline{\phi}_4 = \underline{\phi}_3^*,$$

$$\underline{\phi}_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \underline{\phi}_6 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Im allgemeinen wird nun entsprechend den verschiedenen Hauptdiagonalelementen von $\underline{K}$ die Aussteuerung der Zustandsvariablen unterschiedlich sein. Dies wird im nächsten Schritt korrigiert.

Schritt 3:

Wir transformieren das System mit einer orthogonalen Matrix $\underline{U}$ derart, daß die Hauptdiagonalelemente von $\underline{K}$ gleich werden (siehe Anhang A3, Abschnitt 1.2).

Das so ermittelte Filter besitzt eine normale $\underline{A}$ -Matrix. Die L_2 -Normen aller Zustandsvariablen sind

$$||F_v||_2 = 1/\delta, \quad v = 1(1)n.$$

Das durch Rundung der Zustandsvariablen verursachte Fehlergeräusch am Filterausgang ist minimal und läßt sich nach (A4-20) errechnen.

2. Entwurf eines Filterblockes zweiten Grades mit konjugiert komplexen Polstellen

Im Fall konjugiert komplexer Polstellen $\lambda_{1,2} = \rho e^{\pm j\psi}$ ist nach Abschnitt 3.2.2, Glg. (3-12), die $\underline{A}$ -Matrix von der Form

$$\underline{A} = \rho \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}. \quad (\text{A4-22})$$

Die orthonormalen Eigenvektoren lauten

$$\underline{\phi}_1 = \sqrt{\frac{T}{2}} \begin{bmatrix} 1 \\ j \end{bmatrix} \cdot e^{j\theta}; \quad \underline{\phi}_2 = \sqrt{\frac{T}{2}} \begin{bmatrix} 1 \\ -j \end{bmatrix} \cdot e^{-j\theta} = \underline{\phi}_1^* \quad (\text{A4-23})$$

wobei θ ein beliebiger Phasenwinkel ist.

Der System-Vektor $\underline{b}$ ist dann nach (A4-4) und mit $\beta_2 = \beta_1^*$

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \beta_1 \underline{\phi}_1 + \beta_1^* \underline{\phi}_1^* = \sqrt{2} \begin{bmatrix} \operatorname{Re}\{\beta_1 e^{j\theta}\} \\ -\operatorname{Im}\{\beta_1 e^{j\theta}\} \end{bmatrix}. \quad (\text{A4-24})$$

Die notwendige Skalierungsbedingung (A4-14) wird hier zu

$$|\beta_1|^2 = [1 - \rho^2] \frac{1}{\delta^2} \quad (\text{A4-25a})$$

$$\text{bzw.} \quad b_1^2 + b_2^2 = 2 [1 - \rho^2] \frac{1}{\delta^2}. \quad (\text{A4-25b})$$

Um eine zusätzliche Bedingung für die gleichmäßige Aussteuerung zu erhalten, bestimmt man die Hauptdiagonalelemente der Kovarianzmatrix $\underline{K}$, die sich gemäß (A4-12) ergibt. Die Forderung $K_{11} = K_{22}$ liefert nach Zwischenrechnung die Beziehung /14/

$$\frac{b_1}{b_2} = \frac{\rho^2 \sin(2\psi) \pm \sqrt{1 + \rho^4 - 2\rho^2 \cos(2\psi)}}{1 - \rho^2 \cos(2\psi)}. \quad (\text{A4-26})$$

Nachdem $\underline{b}$ mit (A4-25b) und (A4-26) bestimmt wurde, läßt sich nun $\underline{c}^T = [c_1, c_2]$ ermitteln. Mit (A4-4), (A4-8) und (A4-23) folgt

$$\begin{aligned} \underline{c}^T &= \gamma_1 \underline{\phi}_1^{*T} + \gamma_2 \underline{\phi}_2^{*T} = \frac{R_1}{\beta_1} \underline{\phi}_1^{*T} + \frac{R_1^*}{\beta_1^*} \underline{\phi}_1^T \\ &= \sqrt{2} \left[\operatorname{Re}\left\{ \frac{R_1}{\beta_1 e^{j\theta}} \right\}, \operatorname{Im}\left\{ \frac{R_1}{\beta_1 e^{j\theta}} \right\} \right] \end{aligned} \quad (\text{A4-27})$$

Nach (A4-24) gilt

$$\beta_1 e^{j\theta} = \frac{1}{\sqrt{2}} [b_1 - j b_2]$$

und es folgt nach kurzer Zwischenrechnung

$$c_1 + j c_2 = 2 \frac{R_1 [b_1 + j b_2]}{b_1^2 + b_2^2} \stackrel{\uparrow}{=} \frac{R_1 [b_1 + j b_2]}{(1-\rho^2)} \delta^2 \quad (\text{A4-28})$$

(A4-25b)

Damit sind die Systemmatrizen $\underline{A}$, $\underline{b}$, $\underline{c}^T$ aufgrund der Forderung einer normalen $\underline{A}$ -Matrix und der Skalierungsbedingung eindeutig festgelegt. Es verbleiben keine zusätzlichen Freiheitsgrade zur Minimierung des Rundungsrauschens. Wie man aber leicht zeigt, ist die Optimierungsbedingung (A4-19) hier automatisch erfüllt, denn

$$\begin{aligned} \frac{\delta^2}{n} |R_1| \sum_{v=1}^2 \frac{|R_v|}{1-|\lambda_v|^2} &= \frac{\delta^2}{2} |R_1| \left[\frac{|R_1|}{1-\rho^2} + \frac{|R_1^*|}{1-\rho^2} \right] \\ &= \delta^2 \frac{|R_1|^2}{1-\rho^2} \stackrel{\uparrow}{=} \frac{|R_1|^2}{|\beta_1|^2} = |\gamma_1|^2 = |\gamma_2|^2 \end{aligned}$$

(A4-25a)

Die Ausführungen im vorliegenden Anhang A4 stützten sich im wesentlichen auf die Arbeit von Barnes /14/, sind aber hier in einigen Details genauer ausgeführt.

A N H A N G A 5

Wir betrachten Filter mit normalen $\underline{A}$ -Matrizen der Form

$$\underline{A} = \begin{bmatrix} A_1 & A_2 \\ \bar{A}_2 & A_1 \end{bmatrix}. \quad (\text{A5-1})$$

Die Eigenwerte seien einfach mit $|\lambda_{1,2}| < 1$.

Derartige Matrizen sind kennzeichnend für Normalform-Strukturen. Sie treten aber auch als Zwischenergebnis bei der Bestimmung der rauschoptimalen Zustandsraum-Struktur auf (Abschnitt 3.2.3, Gleichungen (3-21), (3-30)). Im letzteren Fall gewinnt man hieraus das Endresultat durch Transformation mit einer Diagonalmatrix

$$\underline{T} = \delta \cdot \text{Diag} \{ \|F_1\|_2, \|F_2\| \} \quad (\text{L}_2\text{-Skalierung})$$

(siehe Abschnitt 3.2.3, Glg. (3-26)).

Erfolgt die Skalierung nach der L_∞ -Norm, so ist

$$\underline{T} = \text{Diag} \{ \|F_1\|_\infty, \|F_2\|_\infty \}.$$

Das skalierte Filter besitzt die modifizierte $\underline{A}$ -Matrix

$$\underline{A}' = \underline{T}^{-1} \underline{A} \underline{T}.$$

Behauptung:

Für $\lambda_{1,2} = \sigma \pm j\omega$ und $|\lambda_{1,2}| \rightarrow 1$
oder $\lambda_{1,2}$ reell und $|\lambda_1| \rightarrow 1$

gilt

$$\|F_1\|_2 \rightarrow \|F_2\|_2 \quad \text{bzw.} \quad \|F_1\|_\infty \rightarrow \|F_2\|_\infty$$

und damit

$$\underline{T} \rightarrow \delta \cdot \underline{E} \quad \text{bzw.} \quad \underline{T} \rightarrow \underline{E}.$$

Das bedeutet, daß mit $\underline{A}' \rightarrow \underline{A}$ das skalierte System ebenfalls eine normale $\underline{A}$ -Matrix besitzt.

Beweis:

1. Bestimmung der L_2 -Norm

Bezeichnet $F_v(z)$ die Übertragungsfunktion vom Filtereingang zur v -ten Zustandsgröße, dann gilt mit Anhang A3, Glg. (A3-9):

$$\begin{aligned} \|F_v\|_2^2 &= \frac{1}{2\pi} \int_0^{2\pi} |F_v(e^{j\Omega})|^2 d\Omega \\ &= \frac{1}{2\pi j} \oint_{\Gamma} F_v(z) F_v\left(\frac{1}{z}\right) \frac{dz}{z}, \quad v = 1, 2. \end{aligned} \quad (\text{A5-2})$$

Die Integration erfolgt über den Einheitskreis Γ .

Die Partialbruchzerlegung von $F_v(z)$ sei

$$F_v(z) = \sum_{\mu=1}^2 \frac{R_{v\mu}}{z - \lambda_{\mu}}, \quad R_{v\mu} \neq 0. \quad (\text{A5-3})$$

Dann gilt

$$\|F_v\|_2^2 = \frac{1}{2\pi j} \oint_{\Gamma} \left(\sum_{\mu=1}^2 \sum_{\kappa=1}^2 \frac{R_{v\mu}}{z - \lambda_{\mu}} \frac{R_{v\kappa}}{1 - z\lambda_{\kappa}} \right) dz$$

Die Integration liefert das Ergebnis

$$\|F_v\|_2^2 = \sum_{\mu=1}^2 R_{v\mu} \sum_{\kappa=1}^2 \frac{R_{v\kappa}}{1 - \lambda_{\mu} \cdot \lambda_{\kappa}}, \quad v = 1, 2. \quad (\text{A5-4})$$

Fall 1: $\lambda_{1,2} = \sigma \pm j\omega$, $|\lambda_{1,2}| \rightarrow 1$

Unter Berücksichtigung von $\lambda_2 = \lambda_1^*$ und $R_{v2} = R_{v1}^*$ gewinnt man aus (A5-4)

$$\|F_v\|_2^2 = 2 \frac{|R_{v1}|^2}{1 - |\lambda_1|^2} + 2 \operatorname{Re} \left\{ \frac{R_{v1}^2}{1 - \lambda_1} \right\}$$

Für $|\lambda_1| \rightarrow 1$ ist $\operatorname{Re} \left\{ \frac{R_{v1}^2}{1 - \lambda_1} \right\} < \frac{|R_{v1}|^2}{1 - |\lambda_1|^2}$ und es folgt

$$\|F_v\|_2^2 \rightarrow 2 \frac{|R_{v1}|^2}{1 - |\lambda_1|^2}, \quad v = 1, 2. \quad (\text{A5-5})$$

Fall 2: $\lambda_{1,2}$ reell, $|\lambda_1| \rightarrow 1$

$$||F_v||_2^2 \rightarrow \frac{R_{v1}^2}{1-\lambda_1^2}, \quad v = 1, 2. \quad (A5-6)$$

2. Bestimmung der L_∞ -Norm

Nach Anhang A3, Glg. (A3-8) gilt

$$||F_v||_\infty = \max_{\Omega} \{ |F(e^{j\Omega})| \}. \quad (A5-7)$$

Fall 1: $\lambda_{1,2} = \sigma \pm j\omega = \rho e^{\pm j\psi}$, $|\lambda_{1,2}| = \rho \rightarrow 1$

Mit der Partialzerlegung (A5-3) folgt für $\rho \rightarrow 1$

$$||F_v||_\infty \rightarrow \left| \frac{R_{v1}}{e^{j\psi} - \rho e^{j\psi}} \right| = \frac{|R_{v1}|}{1-\rho}, \quad v = 1, 2. \quad (A5-8)$$

Fall 2: $\lambda_{1,2}$ reell mit $|\lambda_1| \rightarrow 1$

$$||F_v||_\infty \rightarrow \frac{|R_{v1}|}{1-|\lambda_1|}, \quad v = 1, 2. \quad (A5-9)$$

Die Beziehungen (A5-5), (A5-6), (A5-8), (A5-9) zeigen, daß für $|\lambda_{1,2}| \rightarrow 1$ oder $|\lambda_1| \rightarrow 1$ die Normen der Zustandsvariablen x_1 und x_2 genau dann gleich sind, wenn R_{11} betragsmäßig gleich R_{21} ist.

3. Bestimmung der Residuen R_{11} und R_{21}

Aus der Zustandsraumbeschreibung

$$\underline{x}(k+1) = \underline{A} \underline{x}(k) + \underline{b} v(k) \quad ; \quad \underline{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

folgt

$$\begin{bmatrix} F_1(z) \\ F_2(z) \end{bmatrix} = \frac{1}{V(z)} \begin{bmatrix} X_1(z) \\ X_2(z) \end{bmatrix} = [z\underline{E} - \underline{A}]^{-1} \underline{b}. \quad (A5-10)$$

Die Auswertung dieser Beziehungen für eine A-Matrix der Form (A5-1) liefert

$$\begin{bmatrix} F_1(z) \\ F_2(z) \end{bmatrix} = \frac{1}{(z-\lambda_1)(z-\lambda_2)} \begin{bmatrix} b_1(z-A_1) + b_2 A_2 \\ -b_1 A_2 + b_2(z-A_1) \end{bmatrix} \quad (\text{A5-11})$$

Fall 1: $\lambda_{1,2} = \sigma \pm j\omega$

Man überprüft leicht, daß

$$\underline{A} = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} \quad (\text{A5-12})$$

ist. Damit geht (A5-11) über in

$$\begin{bmatrix} F_1(z) \\ F_2(z) \end{bmatrix} = \frac{1}{[z-(\sigma+j\omega)][z-(\sigma-j\omega)]} \begin{bmatrix} b_1(z-\sigma) - b_2\omega \\ + b_1\omega + b_2(z-\sigma) \end{bmatrix} \quad (\text{A5-13})$$

Hieraus errechnet man

$$R_{11} = \frac{b_1 - jb_2}{2} \quad ; \quad R_{21} = \frac{b_1 - jb_2}{-2j}$$

bzw.

(A5-14)

$$|R_{21}| = |R_{11}|$$

Fall 2: $\lambda_{1,2}$ reell

A besitzt hier die Form

$$\underline{A} = \begin{bmatrix} \frac{\lambda_1 + \lambda_2}{2} & \pm \frac{\lambda_1 - \lambda_2}{2} \\ \pm \frac{\lambda_1 - \lambda_2}{2} & \frac{\lambda_1 + \lambda_2}{2} \end{bmatrix} \quad (\text{A5-15})$$

Mit (A5-11) folgt

$$\begin{bmatrix} F_1(z) \\ F_2(z) \end{bmatrix} = \frac{1}{(z - \frac{\lambda_1 + \lambda_2}{2})(z - \frac{\lambda_1 - \lambda_2}{2})} \begin{bmatrix} b_1(z - \frac{\lambda_1 + \lambda_2}{2}) \pm b_2 \frac{\lambda_1 - \lambda_2}{2} \\ \pm b_1 \frac{\lambda_1 - \lambda_2}{2} + b_2(z - \frac{\lambda_1 + \lambda_2}{2}) \end{bmatrix} \quad (A5-16)$$

und

$$R_{11} = \frac{b_1 \pm b_2}{2} \quad ; \quad R_{21} = \pm R_{11}$$

bzw.

(A5-17)

$$R_{21} = R_{11}$$

Gemäß (A5-14) und (A5-17) gilt also stets $|F_{11}| = |F_{12}|$. Berücksichtigt man dieses Ergebnis bei der Bestimmung der L_2 - und L_∞ -Normen (Gln. (A5-5), (A5-6) und (A5-8), (A5-9), so folgt für die betrachteten Grenzfälle unmittelbar die Behauptung

$$\|F_1\|_2 \rightarrow \|F_2\|_2 \quad \text{und} \quad \|F_1\|_\infty \rightarrow \|F_2\|_\infty.$$

Voraussetzung ist natürlich, daß die Residuen ungleich Null sind, was aber in praktisch interessierenden Fällen stets erfüllt ist. Verschwindende Residuen bedeuten nämlich, daß die entsprechenden Zustandsvariablen nicht steuerbar sind. Bezüglich der Erregung degeneriert das Filter zu einem System erster oder sogar nullter Ordnung.

A N H A N G A 6

ALLGEMEINE EIGENSCHAFTEN DER MATRIX $\hat{A} = \underline{G}_1^{1/2} \underline{A} \underline{G}_1^{-1/2}$

Wir betrachten Systeme, für welche die Matrizen $\underline{P}$ bzw. $\hat{\underline{P}}$ mit

$$\underline{P} = \underline{G}_1 - \underline{A}^T \underline{G}_1 \underline{A} = \underline{G}_1 - \underline{G}_1^{1/2} \hat{\underline{A}}^T \hat{\underline{A}} \underline{G}_1^{1/2}$$

$$\text{bzw. } \hat{\underline{P}} = \underline{G}_1^{-1/2} \underline{P} \underline{G}_1^{-1/2} = \underline{E} - \hat{\underline{A}}^T \hat{\underline{A}} \quad (\text{A6-1})$$

positiv (semi-) definit sind (vgl. Glg. (2-15)). Dies ist äquivalent zu

$$||\underline{A} \underline{x}||_{\underline{G}_1} \leq ||\underline{x}||_{\underline{G}_1} \quad \forall \underline{x} \in \mathbb{R}^n \quad (\text{A6-2})$$

$$\text{bzw. } ||\hat{\underline{A}} \hat{\underline{x}}||_{\underline{E}} \leq ||\hat{\underline{x}}||_{\underline{E}} \quad \forall \hat{\underline{x}} \in \mathbb{R}^n$$

$$\text{mit } \hat{\underline{x}} = \underline{G}_1^{1/2} \underline{x}$$

Weitere äquivalente Aussagen sind:

1. $\hat{\underline{A}}$ ist Untermatrix einer Orthogonalmatrix.

B_e_w_e_i_s_:

$\hat{\underline{P}}$ ist genau dann positiv (semi-) definit, wenn eine reelle Matrix $\underline{P}_O$ mit

$$\hat{\underline{P}} = \hat{\underline{P}}_O^T \hat{\underline{P}}_O \quad (\text{A6-3})$$

existiert. Da nämlich $\hat{\underline{P}}$ gemäß (A6-1) symmetrisch und somit normal ist, gilt die Darstellung

$$\hat{\underline{P}} = \underline{U}^T \hat{\underline{\Lambda}} \underline{U}$$

$\underline{U}$ = Orthogonalmatrix

$\hat{\underline{\Lambda}}$ = Diagonalmatrix der reellen Eigenwerte λ_v der Matrix $\hat{\underline{P}}$

Wegen $\lambda_v \geq 0$ erhält man eine mögliche Matrix $\hat{\underline{P}}_0$ als

$$\hat{\underline{P}}_0 = \underline{A}^{1/2} \underline{U}.$$

Umgekehrt folgt aus (A6-3)

$$\underline{x}^T \hat{\underline{P}} \underline{x} = (\hat{\underline{P}}_0 \underline{x})^T (\hat{\underline{P}}_0 \underline{x}) = ||\underline{P}_0 \underline{x}||_{\underline{E}}^2 \geq 0 \quad \forall \underline{x} \in \mathbb{R}^n;$$

$\hat{\underline{P}}$ ist also positiv (semi-) definit.

Wir bilden nun mit $\hat{\underline{A}}$ und $\hat{\underline{P}}_0$ die neue $(2n \times n)$ Matrix

$$\hat{\underline{S}}_0 := \begin{bmatrix} \hat{\underline{A}} \\ \hat{\underline{P}}_0 \end{bmatrix}. \quad (\text{A6-4})$$

Ihre Spalten sind orthonormal, denn mit (A6-1), (A6-3) gilt

$$\begin{aligned} \hat{\underline{S}}_0^T \hat{\underline{S}}_0 &= \hat{\underline{A}}^T \hat{\underline{A}} + \hat{\underline{P}}_0^T \hat{\underline{P}}_0 \\ &= \hat{\underline{A}}^T \hat{\underline{A}} + \underline{E} - \hat{\underline{A}}^T \hat{\underline{A}} = \underline{E}. \end{aligned} \quad (\text{A6-5})$$

Durch Ergänzen weiterer n orthonormaler Spaltenvektoren der Dimension $2n$ läßt sich somit eine $(2n \times 2n)$ Orthogonalmatrix konstruieren.

Gehen wir umgekehrt davon aus, daß $\hat{\underline{A}}$ Untermatrix einer Orthogonalmatrix ist, so läßt sich immer ein $\hat{\underline{S}}_0$ gemäß (A6-4) mit der Eigenschaft (A6-5) angeben. Die Zeilenzahl von $\hat{\underline{P}}_0$ muß jetzt nicht notwendigerweise gleich n sein. Mit $\hat{\underline{P}} = \hat{\underline{P}}_0^T \hat{\underline{P}}_0$ folgt aus (A6-5)

$$\hat{\underline{P}} = \hat{\underline{P}}_0^T \hat{\underline{P}}_0 = \underline{E} - \hat{\underline{A}}^T \hat{\underline{A}}.$$

$\hat{\underline{P}}$ ist offenbar positiv (semi-) definit.

q.e.d.

2. $\hat{\underline{A}}$ läßt sich, sofern keine Eigenwerte $\lambda_v = 0$ auftreten, als Produkt einer Orthogonal- und einer Normalmatrix, deren reelle Eigenwerte betragsmäßig kleiner oder gleich 1 sind, darstellen:

$$\hat{\underline{A}} = \underline{U} \tilde{\underline{A}} \quad (\text{A6-6})$$

$\underline{U}$ = (n×n) Orthogonalmatrix

$\tilde{\underline{A}}$ = (n×n) Normalmatrix

Eigenwerte $|\tilde{\lambda}_v| \leq 1$, $\tilde{\lambda}_v$ reell

B e w e i s: (siehe auch /54/)

Setzen wir zunächst die Gültigkeit von (A6-6) voraus, so folgt (Abschnitt 3.2.2)

$$||\hat{\underline{A}} \hat{\underline{x}}||_{\underline{E}}^2 = \hat{\underline{x}}^T \hat{\underline{A}}^T \hat{\underline{A}} \hat{\underline{x}} = \hat{\underline{x}}^T \hat{\underline{A}}^T \tilde{\underline{A}} \hat{\underline{x}} \leq \max_v \{\kappa_v^2\} ||\hat{\underline{x}}||_{\underline{E}}^2 \quad \forall \hat{\underline{x}} \in \mathbb{R}^n \quad (\text{A6-7})$$

wobei κ_v^2 die positiven Eigenwerte von $\hat{\underline{A}}^T \hat{\underline{A}}$ bzw. $\tilde{\underline{A}}^T \tilde{\underline{A}}$ bezeichnet. Da $\tilde{\underline{A}}$ normal ist, gilt aber $\kappa_v^2 = |\tilde{\lambda}_v|^2 \leq 1$ und es folgt (A6-2).

Umgekehrt lassen sich bei Gültigkeit von (A6-2) stets zwei Matrizen $\underline{U}$ und $\tilde{\underline{A}}$ gemäß (A6-6) angeben. Dazu schreiben wir zunächst die symmetrische positiv definite Matrix $\hat{\underline{A}}^T \hat{\underline{A}}$ in der Form

$$\hat{\underline{A}}^T \hat{\underline{A}} = \underline{U}_0 \tilde{\underline{A}}^2 \underline{U}_0^T \quad (\text{A6-8})$$

$\underline{U}_0$ = Orthogonalmatrix

$\tilde{\underline{A}}^2$ = Diagonalmatrix der positiven Eigenwerte $\tilde{\lambda}_v^2$ von $\hat{\underline{A}}^T \hat{\underline{A}}$; $\tilde{\lambda}_v^2 \neq 0$, $v = 1(1)n$, sei hier vorausgesetzt (siehe Bemerkungen am Ende des Beweises)

Wir definieren nun

$$\underline{U} := \hat{\underline{A}} \underline{U}_0 \tilde{\underline{A}}^{-1} \underline{U}_0^T \quad (\text{A6-9})$$

$$\tilde{\underline{A}} := \underline{U}_0 \tilde{\underline{A}} \underline{U}_0^T \quad (\text{A6-10})$$

Damit folgt

$$\underline{U} \tilde{\underline{A}} = \hat{\underline{A}} \quad (\text{A6-11})$$

$\underline{U}$ ist orthogonal, denn

$$\begin{aligned}\underline{U}^T \underline{U} &= \underline{U}_0 \tilde{\underline{A}}^{-1} \underline{U}_0^T \underbrace{\hat{\underline{A}}^T \hat{\underline{A}}}_{(A6-8)} \underline{U}_0 \tilde{\underline{A}}^{-1} \underline{U}_0^T \\ &= \underline{U}_0 \tilde{\underline{A}}^{-1} \tilde{\underline{A}}^2 \underline{U}_0^T \\ &= \underline{U}_0 \tilde{\underline{A}}^{-1} \tilde{\underline{A}}^2 \tilde{\underline{A}}^{-1} \underline{U}_0^T = \underline{E}\end{aligned}$$

$\tilde{\underline{A}}$ ist normal mit den reellen Eigenwerten $\tilde{\lambda}_v$ (A6-10).

Somit ist

$$||\hat{\underline{A}} \hat{\underline{x}}||_{\underline{E}} = ||\tilde{\underline{A}} \hat{\underline{x}}||_{\underline{E}} \leq \max_v(|\tilde{\lambda}_v|) \cdot ||\hat{\underline{x}}||_{\underline{E}} \quad \forall \hat{\underline{x}} \in \mathbb{R}^n.$$

Ferner gilt wegen der Voraussetzung (6-2)

$$||\hat{\underline{A}} \hat{\underline{x}}||_{\underline{E}} \leq ||\hat{\underline{x}}||_{\underline{E}} \quad \forall \hat{\underline{x}} \in \mathbb{R}^n,$$

und es folgt schließlich $|\tilde{\lambda}_v| \leq 1, v = 1(1)n$.

Einschränkung:

In (A6-8) wurde $\tilde{\lambda}_v^2 \neq 0, v = 1(1)n$, vorausgesetzt. Existiert dagegen ein $\tilde{\lambda}_\mu^2 = 0$, mit dem zugehörigen reellen Eigenvektor $\tilde{\underline{x}}_\mu$, so gilt

$$\hat{\underline{A}}^T \hat{\underline{A}} \tilde{\underline{x}}_\mu = \tilde{\lambda}_\mu^2 \tilde{\underline{x}}_\mu = \underline{0}$$

bzw.
$$\tilde{\underline{x}}_\mu^T \hat{\underline{A}}^T \hat{\underline{A}} \tilde{\underline{x}}_\mu = 0.$$

Das bedeutet, daß der ZV-Vektor in einem Filterschritt gleich Null gemacht werden kann. Dazu muß die Matrix $\hat{\underline{A}}$ bzw. $\underline{A}$ aber selbst einen Eigenwert Null aufweisen. Das System ist in diesem Fall nichtrekursiv oder enthält einen nichtrekursiven Anteil.

A N H A N G A 7

Vor: $||\underline{\tilde{x}}(k+1)||_{\underline{G}} \leq \xi ||\underline{x}(k)||_{\underline{G}}$

Beh: $\left| \sum_{\mu=1}^n G_{\mu} \tilde{x}_{\mu}(k+1) \right| \leq \xi ||\underline{x}(k)||_{\underline{G}} \sqrt{\sum_{\mu=1}^n G_{\mu}}$

B_e_w_e_i_s_s:

Zur Vereinfachung der Schreibweise ersetzen wir im folgenden die Vektorkomponenten $\tilde{x}_{\mu}(k+1)$ und $x_{\mu}(k)$ durch $\tilde{x}_{\mu}$ bzw. x_{μ} .

Zunächst gilt

$$\begin{aligned} \left[\sum_{\mu=1}^n G_{\mu} \tilde{x}_{\mu} \right]^2 &= \sum_{\mu=1}^n \sum_{v=1}^n G_{\mu} G_v \tilde{x}_{\mu} \tilde{x}_v \\ &= \sum_{\mu=1}^n G_{\mu}^2 \tilde{x}_{\mu}^2 + \sum_{\mu=1}^n \sum_{\substack{v=1 \\ v \neq \mu}}^n G_{\mu} G_v \tilde{x}_{\mu} \tilde{x}_v \end{aligned} \quad (A7-1)$$

Andererseits findet man

$$\begin{aligned} ||\underline{\tilde{x}}(k+1)||_{\underline{G}}^2 \sum_{v=1}^n G_v &= \left[\sum_{\mu=1}^n G_{\mu} \tilde{x}_{\mu} \right] \left[\sum_{v=1}^n G_v \right] = \\ &= \sum_{\mu=1}^n G_{\mu}^2 \tilde{x}_{\mu}^2 + \sum_{\mu=1}^n \sum_{\substack{v=1 \\ v \neq \mu}}^n G_{\mu} G_v \tilde{x}_{\mu}^2 \end{aligned} \quad (A7-2)$$

Durch Vergleich mit (A7-1) ergibt sich

$$\begin{aligned} \left[\sum_{\mu=1}^n G_{\mu} \tilde{x}_{\mu} \right]^2 &= ||\underline{\tilde{x}}(k+1)||_{\underline{G}}^2 \sum_{v=1}^n G_v - \\ &- \left\{ \sum_{\mu=1}^n \sum_{\substack{v=1 \\ v \neq \mu}}^n G_{\mu} G_v \tilde{x}_{\mu}^2 - \sum_{\mu=1}^n \sum_{\substack{v=1 \\ v \neq \mu}}^n G_{\mu} G_v \tilde{x}_{\mu} \tilde{x}_v \right\} \\ &= ||\underline{x}(k+1)||_{\underline{G}}^2 \sum_{v=1}^n G_v - \sum_{\mu=1}^n \sum_{v=\mu+1}^n G_{\mu} G_v [\tilde{x}_{\mu}^2 - 2\tilde{x}_{\mu} \tilde{x}_v + \tilde{x}_v^2] \end{aligned}$$

$$= ||\tilde{\underline{x}}(k+1)||_{\underline{G}}^2 - \sum_{\mu=1}^n \sum_{v=\mu+1}^n G_{\mu} G_v [\tilde{x}_{\mu} - \tilde{x}_v]^2$$

$$\leq ||\tilde{\underline{x}}(k+1)||_{\underline{G}}^2 - \sum_{v=1}^n G_v$$

Nach Wurzelziehen und Beachtung der Voraussetzung $||\tilde{\underline{x}}(k+1)||_{\underline{G}} \leq \xi ||\underline{x}(k)||_{\underline{G}}$ folgt

$$\left| \sum_{\mu=1}^n G_{\mu} \tilde{x}_{\mu} \right| \leq \xi ||\underline{x}(k)||_{\underline{G}} \sqrt{\sum_{v=1}^n G_v}$$

q.e.d.

L I T E R A T U R

- /1/ L.R. Rabiner and B. Gold: Theory and Applications of Digital Signal Processing. Englewood Cliffs, N.J., Prentice-Hall, 1975
- /2/ A. Fettweis: Digital filter structures related to classical filter networks. AEU 25 (1971), S. 79 - 89
- /3/ W.L. Mills, C.T. Mullis and R.A. Roberts: Digital filter realizations without overflow limit cycles. IEEE Trans. Acoust., Speech, Signal Proces., vol. ASSP-26, Aug. 1978, pp. 334 - 338
- /4/ L.E. Turner and L.T. Bruton: Elimination of granularity and overflow limit cycles in minimum norm recursive digital filters. IEEE Trans. Circuits Syst., vol. CAS-27, Jan. 1980, pp. 50 - 53
- /5/ G.F. Dehner: Ein Beitrag zum rechnergestützten Entwurf rekursiver digitaler Filter minimalen Aufwandes. Ausgewählte Arbeiten über Nachrichtensysteme Nr. 23, hrsg. von H.W. Schüßler, Univ. Erlangen-Nürnberg, 1976
- /6/ H.W. Schüßler: Digitale Systeme zur Signalverarbeitung. Springer-Verlag, Berlin 1973
- /7/ T.A.C.M. Claasen and L.O.G. Kristiansson: Necessary and sufficient conditions for the absence of overflow phenomena in a second-order recursive digital filter. IEEE Trans. Acoust., Speech, Signal Proc., ASSP-23, Dec. 1975, pp. 509 - 515
- /8/ T.A.C.M. Claasen, W.F.G. Mecklenbräuker and J.B.H. Peek: On the stability of the forced response of digital filters with overflow nonlinearities. IEEE Trans. Circuits and Systems, CAS-22, Aug. 1975, pp. 692 - 696

- /9/ K. Meerkötter: Beiträge zur Theorie der Wellendigitalfilter. Dissertation, Ruhr-Univ. Bochum, 1979
- /10/ B. Eckhardt, W. Winkelkemper: Implementation of a second order digital filter section with stable overflow behaviour. NTZ, Bd. 26, 1973, S. 282 - 284
- /11/ L.B. Jackson: On the interaction of roundoff noise and dynamic range in digital Filters. Bell Syst. Techn. J., Bd. 49, 1970, S. 159 - 184
- /12/ P.M. Ebert, J.E. Mazo, M.G. Taylor: Overflow oscillations in digital filters. Bell Syst. Techn. J., Bd. 48, 1969, S. 2999 - 3020
- /13/ C.W. Barnes and A.T. Fam: Minimum norm recursive digital filters that are free of overflow limit cycles. IEEE Trans. Circuits and Systems, CAS-24, Oct. 1977, S. 569 - 574
- /14/ C.W. Barnes: Roundoff noise and overflow in normal digital filters. IEEE Trans. Circuits and Systems, CAS-26, March 1979, S. 154 - 159
- /15/ C.W. Barnes, S. Shinnaka: Stability domains for second-order recursive digital filter in normal form with "matrix power" feedback. IEEE Trans. Circuits and Systems, CAS-27, Sept. 1980, S. 841 - 843
- /16/ C.W. Barnes, S.-H. Leung: Use of transversal-recursive structures for efficient realization of low-noise digital filters with decimated output. IEEE Trans. Acoust., Speech, Signal Proc., ASSP-28, Dec. 1980, S. 645 - 651
- /17/ C.W. Barnes: Error feedback in normal realizations of recursive digital filters. IEEE Trans. Circuits and Systems, CAS-28, Jan. 1981, S. 72 - 75

- /18/ C.W. Barnes: Intuitive explanation for low roundoff noise in normal realizations of narrow-band digital filters. IEEE Trans. Acoust., Speech, Signal Proc., ASSP-29, April 1981, S. 311 - 312
- /19/ C.W. Barnes and T. Miyawaki: Roundoff noise invariants in normal digital filters. IEEE Trans. Circuits and Systems, CAS-29, April 1982, S. 251 - 256
- /20/ R. Zurmühl: Matrizen. Springer-Verlag, Berlin, 1964
- /21/ C.T. Mullis and R.A. Roberts: Synthesis of minimum round-off noise fixed point digital filters. IEEE Trans. Circuits and Systems, CAS-23, Sept. 1976, S. 551 - 562
- /22/ S.Y. Hwang: Minimum uncorrelated unit noise in state space digital filtering. IEEE Trans. Acoust., Speech, Signal Proc., ASSP-25, Aug. 1977, S. 273 - 281
- /23/ L.B. Jackson, A.G. Lindgren, Y. Kim: Optimal synthesis of second-order state-space structures for digital filters. IEEE Trans. Circuits and Systems, CAS-26, March 1979, S. 149 - 153
- /24/ R. Czarnach: Entwurf und Aufbau eines Standardbausteins für rekursive digitale Filter. Diplomarbeit am Institut für Nachrichtentechnik der Universität Erlangen-Nürnberg, Erlangen 1978
- /25/ C.T. Mullis, R.A. Roberts: Roundoff noise in digital filters: frequency transformations and invariants. IEEE Trans. Acoust., Speech, Signal Proc., ASSP-24, Dec. 76, S. 538 - 550
- /26/ L.B. Jackson: Limit Cycles in state-space structures for digital filters. IEEE Trans. Circuits and Systems, CAS-26, Jan. 1979, S. 67-70
- /27/ A. Sedlmeyer: Structures for wave digital filters and their time realization on a minicomputer. Dissertation Ruhr-Univ. Bochum, 1974

- /28/ H. Scheuermann: Grundlagen der Wellendigitalfilter. Wiss. Ber. AEG-TELEFUNKEN 54 (1981), 1-2, S. 11-24
- /29/ K. Meerkötter, W. Wegener: A new second-order digital filter without parasitic oscillation. AEU, Bd. 29, 1975, S. 312 - 314
- /30/ W. Kühnel: Aufwandsuntersuchungen bei Wellendigitalfiltern. Diplomarbeit am Institut für Nachrichtentechnik der Universität Erlangen-Nürnberg, Erlangen 1974
- /31/ A.H. Gray, Jr., and J.D. Markel: Digital lattice and ladder filter synthesis. IEEE Trans. Audio, Electroacoust., AU-21, Dec. 1973, S. 491 - 500
- /32/ A.H. Gray, Jr., and J.D. Markel: A normalized digital filter structure. IEEE Trans. Acoust., Speech, Signal Processing, ASSP-23, June 1975, S. 268 - 277
- /33/ J.D. Markel and A.H. Gray, Jr.: Roundoff noise characteristics of a class of orthogonal polynomial structures. IEEE Trans. Acoust., Speech, Signal Proc., ASSP-23, Oct. 1975, S. 473 - 486
- /34/ J.D. Markel and A.H. Gray, Jr.: Fixed-point implementation algorithms for a class of orthogonal polynomial filter structures. IEEE Trans. Acoust., Speech, Signal Proc., ASSP-23, Oct. 1975, S. 486 - 494
- /35/ A.H. Gray, Jr.: Passive cascaded lattice digital filters. IEEE Trans. Circuits and Systems, CAS-27, May 1980, S. 337 - 344
- /36/ K.-P. Nürnberg: Quantisierungseffekte bei rekursiven digitalen Filtern in Leiterstruktur. Diplomarbeit am Institut für Nachrichtentechnik der Universität Erlangen-Nürnberg, Erlangen 1982

- /37/ W.I. Smirnow: Lehrgang der höheren Mathematik, Teil I. VEB Deutscher Verlag der Wissenschaften, Berlin 1967, S. 337 ff.
- /38/ H. Boseck: Einführung in die Theorie der linearen Vektorräume. VEB Deutscher Verlag der Wissenschaften, Berlin, 1965, S. 49
- /39/ H. Wellmann: Quantisierungseffekte bei Wellendigitalfiltern. Diplomarbeit am Institut für Nachrichtentechnik der Universität Erlangen-Nürnberg, Erlangen 1983
- /40/ J. Zeman, A.G. Lindgren: Fast digital filters with low round-off noise. IEEE Trans. Circuits and Systems, CAS-28, July 1981, S. 716-723
- /41/ J. Zeman, A.G. Lindgren: Fast state-space decimator with very low roundoff noise. Signal Processing, Vol. 3, October 81, S. 377 - 388
- /42/ C.W. Barnes, S. Shinnaka: Finite word effects in blockstate realizations of fixed-point digital filters. IEEE Trans. Circuits and Systems, CAS-27, May 1980, S. 345 - 349
- /43/ C.W. Barnes, S. Shinnaka: Block-shift invariance and block implementation of discrete-time filters. IEEE Trans. Circuits and Systems, CAS-27, Aug. 1980, S. 667 - 672
- /44/ H.J. Kolb: Der Mikrosignalprozessor MSP 82. Ausgewählte Arbeiten über Nachrichtensysteme Nr. 51, hrsg. von H.W. Schüßler, Univ. Erlangen-Nürnberg, 1982
- /45/ W. Zahnleiter: Rekursive digitale Filterung mit Hilfe eines vorhandenen Mikrosignalprozessors. Diplomarbeit am Lehrstuhl für Nachrichtentechnik der Universität Erlangen-Nürnberg, Erlangen 1981

- /46/ R. Czarnach, R. Rabenstein, H.W. Schüßler: On the stability of certain time-varying digital filters. Proc. of the Second European Signal Processing Conference, EUSIPCO-83, Erlangen, 1983, S. 77 - 80
- /47/ S. Horvath, Y.S. Wu: On the stability of linear time-varying and adaptive rekursive IIR-filters. Proc. of the International Conference on Digital Signal Processing, Florence, 1981, S. 18 - 22
- /48/ D.M. Wiberg: State Space and linear Systems. Schaum's Outline Series, McGraw Hill Book Company, 1971
- /49/ R. Unbehauen: Systemtheorie: Eine Darstellung für Ingenieure. R. Oldenbourg Verlag, München - Wien, 3. Auflage, 1980
- /50/ R.A. Meyer and C.S. Burrus: A unified analysis of multirate and periodically time-varying digital filters. IEEE Trans. on Circuits and Systems, Vol. CAS-26, no. 3, March 1979, S. 154 - 159
- /51/ M. Büttner: Untersuchungen über Grenzzyklen in digitalen Filtern. Dissertation Univ. Erlangen-Nürnberg, Erlangen 1977
- /52/ H.W. Schüßler: Netzwerke, Signale und Systeme, Band II, Springer-Verlag, Berlin 1984
- /53/ D. Mitra: Large amplitude, self-sustained oscillations in difference equations which describe digital filter sections using saturation arithmetic. IEEE Trans. Acoustics, Speech and Signal Process., Vol. ASSP-25, Apr. 1977, S. 134 - 143
- /54/ F.R. Gantmacher: Matrizenrechnung I. VEB Deutscher Verlag der Wissenschaften, Berlin 1965, S. 263 ff

Lebenslauf

=====

Roland Czarnach

28.6.1951 geboren in Hersbruck (Pegnitz)
Eltern: Roman und Elfriede Czarnach
Staatsangehörigkeit: deutsch

1957 - 1961 Volksschule in Velden und Neuhaus/Pegnitz

1961 - 1970 math.-naturw. Gymnasium in Pegnitz
1970 Abitur

1970 - 1972 Bundeswehرداری

1972 - 1978 Studium der Elektrotechnik an der Universität
Erlangen - Nürnberg
Fachpraktikum bei den Firmen
Siemens AG, Forschungszentrum, Erlangen
Kraftwerk Union, Rechenzentrum, Erlangen
1978 Abschluß des Studiums mit Diplom

1978 - 1983 wissenschaftlicher Mitarbeiter am Lehrstuhl für
Nachrichtentechnik der Universität Erlangen -
Nürnberg

seit 1983 Entwicklungs-Ingenieur bei der Firma Gossen,
Erlangen

Roland Czarnach

Druck: Copy Center 2000
Obere Karlstraße 30
8520 Erlangen